Marymount University
College of Business, Innovation, Leadership, and Technology (BILT)

Dissertation Approval Form

We hereby approve the dissertation of

Michael Williams

Candidate for the Doctor of Science in Cybersecurity

Date: 19 Dec 2022

Signature: __Andrew Hall__
Chair: __Dr. Andrew Hall__

Date: 12/27/2022

Signature: __Nathan Green__
Member: __Dr. Nathan Green__

Date: 1/12/2023

Signature: __Arimeta Brown__
Member: __Dr. Arimeta Brown__

Signature: __________________
External Reviewer: _____________

Accepted:

Jonathan Aberman, LLM
Dean, College of Business, Innovation, Leadership and Technology
DETECTING INSIDER THREATS USING MACHINE LEARNING AND TRUST SCORES

By
Michael Williams

Andrew Hall, Ph.D., Committee Chair
Nathaniel Green, Ph.D., Committee Member
Armeta Brown, DCS, External Reader

Jonathan Aberman, LLM, Dean
College of Business, Innovation, Leadership, and Technology

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Science
May 2022
Acknowledgments

I express my heartfelt gratitude to all my professors and their support in my studies. I recognize my external committee member for dedicating their time to help in enhancing my dissertation. I thank my committee members for their expertise and guidance. Finally, I thank my friends and fellow doctoral classmates for our shared knowledge.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>5</td>
</tr>
<tr>
<td>Abstract</td>
<td>7</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>9</td>
</tr>
<tr>
<td>Research Problem</td>
<td>15</td>
</tr>
<tr>
<td>Statement of Purpose</td>
<td>19</td>
</tr>
<tr>
<td>Research Questions</td>
<td>19</td>
</tr>
<tr>
<td>Assumptions</td>
<td>20</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>21</td>
</tr>
<tr>
<td>Limitations of Study</td>
<td>22</td>
</tr>
<tr>
<td>Dissertation Chapters and Organization</td>
<td>22</td>
</tr>
<tr>
<td>Definitions and Terms</td>
<td>25</td>
</tr>
<tr>
<td>Chapter 2: Literature Review</td>
<td>26</td>
</tr>
<tr>
<td>Background</td>
<td>26</td>
</tr>
<tr>
<td>Literature Search Strategy</td>
<td>28</td>
</tr>
<tr>
<td>Contemporary Context</td>
<td>30</td>
</tr>
<tr>
<td>Historical Background</td>
<td>31</td>
</tr>
<tr>
<td>Review of the Literature</td>
<td>33</td>
</tr>
<tr>
<td>Kandias et al. (2013)</td>
<td>33</td>
</tr>
<tr>
<td>Alahmadi et al.</td>
<td>33</td>
</tr>
<tr>
<td>Kelley et al.</td>
<td>34</td>
</tr>
<tr>
<td>Fujii et al.</td>
<td>36</td>
</tr>
<tr>
<td>Park et al.</td>
<td>36</td>
</tr>
<tr>
<td>Summary</td>
<td>37</td>
</tr>
<tr>
<td>Chapter 3: Methodology</td>
<td>39</td>
</tr>
<tr>
<td>Rationale for the Research Approach</td>
<td>39</td>
</tr>
<tr>
<td>Population</td>
<td>41</td>
</tr>
<tr>
<td>Sampling and Sampling Procedures</td>
<td>41</td>
</tr>
<tr>
<td>Data Collection (Primary Data)</td>
<td>43</td>
</tr>
<tr>
<td>Data Collection Procedures</td>
<td>44</td>
</tr>
<tr>
<td>Trustworthiness (Validity and Reliability)</td>
<td>77</td>
</tr>
<tr>
<td>Summary</td>
<td>80</td>
</tr>
<tr>
<td>Chapter 4: Data Analysis</td>
<td>81</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>82</td>
</tr>
<tr>
<td>The analyzing of these posts sought to determine whether reactions</td>
<td>89</td>
</tr>
<tr>
<td>could be generalized cross-culturally</td>
<td></td>
</tr>
<tr>
<td>Chapter 5: Discussion, Conclusions, and Recommendations</td>
<td>100</td>
</tr>
<tr>
<td>Interpretation of Findings</td>
<td>103</td>
</tr>
</tbody>
</table>
Abstract

Insider threats are security problems that pose complex challenges for all organizations and are becoming more prevalent. The issue has grown since the invention of social media and the internet. This research focuses on using machine learning and trust scores to identify possible insider threats through posts on Facebook. The study will begin by collecting data from the Peace Corps Facebook page and constructing a trust score that will be used as a baseline score. This score will be compared to the trust scores of posts from CNN and Fox News Facebook social media pages to determine if posts on these pages are insider threat concerns. The posts on Facebook pages were retrieved using Facepager, exported to comma-separated value (CSV) files, and analyzed using R programming and search words to acquire a trust score. The experiment categorized the scores into five groups: extremely positive, above normal, normal, concerning, and extremely concerning. The posts tagged as concerning could be used as a mechanism to investigate the post or the individual behind the post. The research did not collect actual names to identify users but instead assigned identification numbers to each post, so that the individual behind the post remains anonymous.
This dissertation and experimental study will attempt to minimize the possibility of security leaks and breaches in organizations by contributing to the research community. 

*Keywords*: trust scores, machine learning, cybersecurity, R programming, insider threat, Facepager, and comma-separated values (csv).
Chapter 1: Introduction

Bill Fair collaborated with mathematician Earl Isaac to create Fair, Isaac & Company (FICO), and their company’s goal was to create a standardized, impartial credit scoring system (Demyanyk, 2008). They began selling this credit scoring system two years later to American companies. The company FICO today has evolved to become an industry standard. The FICO scoring system falls between 300 and 850, determined by factors encompassing payment history, owed amounts, credit history length, diverse types of credit used, and credit inquiry frequency (Demyanyk, 2008). Fannie Mae and Freddie Mac suggested using the FICO score in mortgage loan lending in 1995 (Demyanyk, 2008). Today, the individual FICO score comes from the three national credit bureaus (Experian, Transunion, Equifax). Today, opponents of social trust scores think FICO and other credit scores may be a “thing of the past” and think of them as just a scoring system. The third season of the famous British television series “Black Mirror” featured a scoring application that allowed people to rate their interactions and share status updates (Jones et al., 2016). The thought of sharing a judgmental score with everyone encountered may seem intrusive. Still, the benefits of allowing people access to this information would help people decide whether they should conduct business (e.g., hire a babysitter, lease renters a home, order food from a delivery driver).
A similar study showing a restaurant's rating included a novel reviews dataset from the website YELP.com (Luca, 2016). Luca (2016) found a restaurant's rating increased by fifty percent if the business obtained at least fifty positive reviews compared to a restaurant with fewer than ten positive reviews. Since YELP users can read comments and ratings, the increased number of reviews could concern rate-sensitive restaurants. There is a negative effect on businesses, people, or products that comes with a viewable text review by the world, and both companies and people feel that the world is judging them. On the other hand, trust scores are derived from a scoring algorithm, and users cannot usually leave a written comment, which could unequally sway reviewers' opinions (Luca, 2016).

Apple has quietly introduced a trust score based on how customers use their iPhones and other related devices. In 2016, these devices were in the hands of 90.1 million Americans (Thao & Tsanthaiwo, 2017). This enormous user base allows the creation of scores by tracking both calls and emails in pursuit of reducing fraud. The score allows or rejects the user's online purchases based on a calculated trust score for each device. According to Apple, these scores live for a fixed time on their servers (Thao & Tsanthaiwo, 2017). Apple is assigning device trust scores based on network information, including the number of phone calls and email metadata. This data helps Apple identify scammers using Apple's services and devices to commit fraudulent activity. Unlike the Chinese trust scores, where other users and companies can share scores and data, Apple has genuinely tried to protect users' data by temporarily storing
the scores on its servers and protecting the data used to arrive at this score (Thao & Tsanthaiwo, 2017).

It is now commonplace for organizations to collect personal information to arrive at a trust score designed to protect business and provide added value to consumers. This collection of data may enable the protection of a country’s greatest assets from insider threats. Many countries are familiar with financial credit ratings, and in recent years some have seen a proliferation of consumer credit rating systems to rate their citizens. These systems consist of online platforms (e.g., eBay, Uber, Airbnb).

Canada uses a credit reporting system most like the United States. They also use the Equifax and Transunion credit bureaus, but the Canadian system has a high score of 900 compared to the U.S. maximum credit score of 850. The United Kingdom’s credit systems are similar to the United States but also include voter registration data, as registering is viewed positively by lenders and improves the credit scores of its citizens (Stephenson, 2020). Unlike the US, Canada, and the UK, many countries do not rely solely on the main three credit bureaus (i.e., Experian, Transunion, Equifax) utilized by these three countries (Stephenson, 2020). For instance, the Netherlands uses a credit system called Krediet Registatie (BKR) to evaluate its borrowers (Curley, 2018). Australia uses the credit bureau Veda (acquired by Equifax), Dun and Bradstreet, Experian, and the Tasmanian Collection Service. India uses a TransUnion partner called the Bureau of Information India (Curley, 2018).
While these countries have adopted a similar credit system to the United States, the People's Bank of China (PBOC), controlled by the Chinese Communists Party (CCP), pushed back on the plans of eight private companies licensed to implement credit score pilot programs in 2015. The CCP did this delay due to concerns about private companies' overreach and, more importantly, the government's plan to develop its credit scoring system (Bach, 2020). The Chinese Social Credit System (SCS) was an unprecedented and logistically challenging effort. Still, the CCP was adamant that the program would create new metrics to connect economic and moral behavior through a “sincerity” index (Bach, 2020, pp. 491). The SCS increased access to credit, decreased trust in the market, leveraged digital technology for population management (Bach, 2020, pp.493). In this system, “sincerity” (chengxin 诚信) translates as social integrity, honesty, or credibility. Sincerity appeared 135 times in the SCS implementation documents (Bach, 2020). The goal of the SCS system can be encapsulated in an oft-quoted phrase within its original planning document: The plan envisions a future where “the trustworthy roam everywhere under heaven while making it hard for the discredited to take a single step” (McSorley, 2021, pp.6).

Originally the word credit in English referred to honor and virtue, but it has become known today as one’s ability to repay debt and is increasingly used in finance (Bach, 2020, pp.491). The term “credit” in Chinese (xinyong 信用) uses the broader meaning of honor and virtue and includes financial credit. Xinyong is more active than its English equivalent since the second character, yong (用), means to use, mobilize, or
deploy, and xin (信) means belief, trust, or trustworthiness (Bach, 2020, pp.491).

Therefore, it makes the concept of "credit" in China well-positioned in both the economic and moral domains. This allows the SCS to function across technical, financial, ethical, and political registers as a part of a modern utopian vision of a meritocratic society (Ortiz, 2013). Social credit has emerged to address instability across productivity, trust, and population control alongside China's surging market reforms during the latter part of the 20th century and early 21st century (Ragnell, 2020).

The SCS assigns a trustworthiness score to Chinese citizens, dependent on society's behavioral systems, and issues immediate rewards and penalties through feedback on the platform. The Chinese social credit score includes data from citizens' social networks and internet activity, consumption, debt and bill payments, legal issues, travel history, public behavior, and marital status (Ragnell, 2020). While Chinese citizens receive luxurious rewards and privileges for abiding by or exceeding standards of trustworthiness, those with low SCS scores experience punishment by limited choices and higher prices (Ragnell, 2020). The SCS is designed to create an underclass of people isolated from privileged society and makes it nearly impossible for them to integrate back into their community (Froidmont et al., 2018).

In emerging economies, there is concern that implementing social credit score systems will alienate a large population segment. There has been a lot of experimentation in the digital finance sector to identify creditworthiness. These experiments concluded
that conventional credit scoring systems emphasize correlation, as causation can be difficult to demonstrate through statistical tests and can introduce unintended discrimination (Mac Síthigh & Siems, 2019).

Although possibly discriminative and alienating sections of society, social media tracking of citizens has been widely adopted and employed across the United States. When a customer logs in to their Starbucks or Airbnb account or makes a reservation on OpenTable, vast amounts of information about them are compiled instantly into a single score. This score, alongside other personal data, is then used to determine the probability that the authenticated entity is either a robot or a risky human (Mims, 2019). Similarly, an employee’s data can be continuously monitored, evaluating the risk they might pose to their employer.

In 2013 Edward Snowden, a Booz Allen Hamilton contractor, copied thousands of classified National Security Agency (NSA) documents without authorization and later leaked those documents to Glenn Greenwald, a reporter from The Guardian (Mokrosinska, 2020). The classified information was subsequently published by media outlets worldwide and made available to friends and adversaries of the United States (Mokrosinska, 2020). If a machine learning algorithm could construct a trust score from publicly available social media data that could assist in identifying insider threats, possibly the risk from insider threat activities could be reduced.

This research explores how trust scores can be created for social media posts by using a group’s social media post data to construct a baseline score. I created a band of trust scores and attempted to identify users that deviate from organizational norms. This
research provides a methodology to focus insider threat prevention efforts and possibly detect insider threats within an organization as they occur.

Lim et al. (2017) proposed an unsupervised machine-learning model that could identify real-world latent infectious diseases by mining social media data. The researcher extracted data from social media, including usernames and temporal and geospatial information. The textual scraping was used with symptom discovery through social media data to identify whether a social media message contains an individual's potential symptoms related to latent infection.

Similarly, this study will use machine learning models and statistical analysis with R to identify social media posts of concern concerning an insider threat. This research associated a trust score for each post collected and compared this to a baseline trust score. The baseline trust score is developed from Peace Corps social media posts, largely consisting of Peace Corps members and used to create our baseline score. Since Peace Corps members are monitored by fellow members and can be reported to management for unethical posting on social media, it makes them a logical group for a baseline score (Peace Corps (US), 2006).

Research Problem

In addressing the rising number of insider threats, this dissertation research has two primary purposes. One purpose consists of using machine learning and statistical programming languages to evaluate users' posts on social media compared against a
baseline score of Peace Corp Facebook page posts. The other is to examine the effects of implementing a trust score to monitor employees and the insider threat problem. As shown in Figure 1, “ransomware” attacks account for nine percent of threat incidents to American companies. Surprisingly with the amount of phishing training available, “phishing” attacks account are the third most common incident type at fifteen percent. “Denial of service” attacks account for ten percent of threat incidents, and “other incidents” attacks account for the remaining nineteen percent (IBM/ObserveIT, 2020). The problem this research will attempt to address is “unauthorized discloser or breaches” attacks accounting for forty-seven percent, which is illustrated in figure 1. Machine learning and trust scores could be used to proactively reduce breaches that fall under unauthorized disclosure consisting of forty-seven percent of insider threats.
These research goals will adopt the theory of trust as the theoretical framework because trustworthiness is an essential element that binds employees with access to sensitive information, leadership, and the public together. Trust allows employees to work together as a cohesive team (Fine & Holyfield, 1996). The understanding that employees must be trustworthy goes well with trust scores being implemented in organizations because, in recent years, businesses have allowed more people who hold sensitive information. Due to cloud services and Google drives, employees bringing their own devices to work have multiple industrial or governmental espionage opportunities. This research will leverage a multi-methodological approach. The first phase will use a
social media scrapping tool (Facepager) to retrieve posts from three social media pages, including posts from our baseline social media page. Since Peace Corps members are held to a higher ethical standard by members in their organization and each member has a responsibility to report other members for unethical behavior, the posts retrieved from the Peace Corp social media site should not contain information suggesting the users are insider threats. However, there could be a few outliers. This dataset will be ideal for constructing a baseline for revealing possible insider threats on social media. The second phase will use R programming to format and analyze each post with a keyword list and existing emoji reactions. Lastly, the experiment will obtain a trust score for each post from the previous analysis and determine if these are concerning by comparing the scores to our baseline dataset score.

Since profile information (education level, gender, marital status, race, or age) will not be collected, the research will not identify respondents of posts. Posts and comments will be solely identified by identification numbers, as Facebook has recently prevented the collection of profile data. The researcher has chosen to analyze comments, posts, and reactions to public posts. Implementation of trust scores to locate insiders on social media may assist businesses to protect sensitive data. Businesses have allowed more people and employees into their secure organization who can collect massive amounts of data using cloud services and Google drives in recent years (Luca, 2016). These people may have legitimate access to an organization’s sensitive data, but the scope for exploitation or exploitation has hugely increased (Luca, 2016).
Statement of Purpose

This quantitative study will test the theory that trust scores developed from Peace Corps Facebook page data can help detect and categorize posts as potential threats by comparing their trust score using statistical methods and machine learning. A low trust score indicates a potential concern when the individual has authorized access to an organization's assets. The trust score will be defined as a metric that is assigned to each post.

Research Questions

The following research questions were addressed:

1. Could a trust score be created using machine learning techniques to reduce insider threats?
2. Could a trust score be used to evaluate the trustworthiness of current and future employees?
3. Could a social media page of members of the Peace Corps be used to create a model that will be useful for detecting insider threats on another Facebook site?
4. What are the security and privacy implications of using social data and a trust score to assess current and prospective employees?

Initially, employees may see surveillance of their activities on social media as a violation of their privacy. However, the benefits of protecting businesses and consumers may outweigh the concerns of confidentiality that technology has well-protected.
According to the Cost of Insider Threats (2020) study, as shown in Figure 2, insider threats cost American companies a total of 800 million dollars in 2016. The cost for American companies was slightly reduced in 2017 to 788 million dollars but increased in 2018 to 890 million. In 2019, insider threat cost reached 977 million. (IBM/ObserveIT, 2020). 2020 had its most significant increase in insider threat cost of 140 million. This brought the total annual cost due to insider threats for American companies to over a billion dollars annually.

Figure 2

Cost of Insider Threats by Year (IBM/ObserveIT, 2020)

Note. Each year within the chart represents the amount in millions which cost the companies collectively for Insider Threats. This cost has increased every year since 2017 with the year of 2020 exceeding a billion dollars.

Assumptions

This study assumes that the period over which the posts were collected and analyzed from each social media site will reflect future time periods. Collecting every post for the three social media pages over a two-month period would have generated an enormous amount of data, and consequently, this experiment sampled 500,000 posts. As private organizations and federal agencies would be interested in protecting their
reputation, brand, and proprietary information, it is assumed that this study will provide a template for future research.

**Significance of the Study**

Shortly after the Snowden incident, the U.S. government began questioning the background investigation process that allowed him to be in a cleared position with access to classified information. After months of investigation on background investigation practices, in February 2014, the committee came out with a report detailing the three major flaws in the background investigation procedures (Mokrosinska, 2020). One flaw was the inability to receive information from law enforcement agencies about employee records continuously. The second was a lack of continuous financial monitoring. The third was privacy barriers and regulations prohibiting background investigators from accessing continuously monitored social media and communication information (Sarah, 2015). These vulnerabilities in the background investigation process might have been partially responsible for the unauthorized disclosure of America’s most sensitive secrets.

Organizations typically use a complementary technology for their systems called anomaly detection for insider security threat detection. Network systems that are expected to remain structured, and controlled using anomaly detection, will have the ability to log fluctuations that deviates from the regular traffic (Yuan, 2021). These anomalies can be stored for each user and system and then compared with user habits using machine learning. While this method may help to detect threats subsequently, it
does not proactively prevent the insider threat from gaining access to sensitive information. This research will use machine learning and trust scores to determine if users posting comments inside a Facebook social media site could be a possible insider threat concern and need further investigation. A trust score methodology used to calculate a score from the user’s social media data will make significant research contributions. While this data is essential in calculating a trust score, exploring the security and privacy implementations of social network data for evaluating current and future employees will contribute to the future body of research. Organizations in the public and private sectors could benefit from suggestions in this research to detect possible insider threats and enhance their background investigation process or replace it altogether.

Limitations of Study

This study will not reveal or store users' names or personal identifying information from social media profiles; due to Facebook not allowing this data to be queried and the inability of the Facepager tool to retrieve this data without permission. The study will not reveal information about an employee’s place of employment for privacy and security reasons. However, this research will assess users' posts and comments on Facebook social media pages and reveal individual posts and comments that could pose a concern compared to users that are not objectively viewed as insider threats.

Dissertation Chapters and Organization

This dissertation is comprised of five chapters. Following the introductory chapter, Chapter 2 provides an overview of the conceptual definition and characteristics
of trust scores and a review of existing research literature about research done on trust scores and machine learning to evaluate insider threats. Chapter 3 will organize the study within a methodological tradition and provide a rationale for the approach describing the research setting and data sample. Chapter 4 will describe unsupervised machine learning used for the data collection and present data analysis results. It will also present the hypotheses based on the theoretical and empirical justifications derived from the literature review. Chapter 5 includes discussions, summarizes conclusions, and details any recommendations. Lastly, it discusses the limitations and suggestions for future research directions.

Machine learning is a subfield of computer science and can be categorized as a process of arriving at a particular result by analyzing data (Burkov, 2019). The subfield can build a statistical model based on one or many datasets, and that model is used to solve practical problems. Learning of this type can be categorized as supervised, semi-supervised, unsupervised, or reinforcement learning (Burkov, 2019).

Supervised learning is a machine learning approach that uses labeled datasets to train or supervise algorithms to classify data or predict outcomes accurately (Burkov, 2019). It tends to be more accurate than unsupervised learning. Using labeled inputs and outputs, the model can measure its accuracy and learn over time (Burkov, 2019). Supervised learning can be separated into two types of problems when dealing with
datasets: classification and regression (Burkov, 2019). Your email provider can use these classification methods to classify spam in a separate folder from your primary inbox.

Regression models are another form of supervised learning that uses an algorithm to understand the relationship between dependent and independent variables. These models help predict numerical values based on data points, such as a business sales revenue projection.

Unsupervised learning depends on machine learning algorithms to analyze and cluster unlabeled data sets (Burkov, 2019). The algorithms uncover hidden patterns in data with no human intervention. While unsupervised learning does not require upfront intervention, the output variables need validation (Burkov, 2019).

Semi-supervised is a happy medium between supervised and unsupervised which is best used when it’s difficult to extract relevant features from data that have a high volume of data (Burkov, 2019). It’s useful for medical images, where a small amount of training data can significantly improve accuracy.
Definitions and Terms

**Comma-separated value (CSV) file:** is a delimited text file that uses separators usually in the form of a comma or semi-colon for exchanging and converting data between various spreadsheet programs for processing.

**Espionage:** the practice of spying or of using spies, typically by governments, to obtain political and military information (Band et al., 2006).

**Trust Score in Social Media (TSM):** measuring individual users' trust levels in a social network (Roy et al., 2017).

**Insider Threat:** the potential for an individual who has or had authorized access to an organization’s assets to use their access, either maliciously or unintentionally, to act in a way that could negatively affect the organization (Park et al., 2018).

**Machine Learning:** is a subfield of computer science that is concerned with building algorithms which, to be useful, rely on a collection of examples of some phenomenon (Burkov, 2019).

**Social Credit System (SCS):** a use of technology which allowed multiple actors to turn human behavior and interactions into a point scoring system on behalf of the Chinese state's goal of modifying the larger social environment (Bach, 2020).
Chapter 2: Literature Review

This chapter describes previous work on trust scores and various techniques used to detect insider threats. Whether it’s inside a Sensitive Compartmented Information Facility (SCIF) used to deny unauthorized personnel or an open access office, insider threats are a major concern in today’s workplace. The following review of literature in this chapter confirms that insider threats present a problem that go beyond attempts at solely deterring at the workplace, discusses specific and general solutions, and concludes that social media data initiatives are needed to protect today's and tomorrow's most sensitive information, people and critical infrastructure.

Background

When the two words “insider threat” are broken down, an insider is defined as “an individual who has special knowledge or access to sensitive, confidential or classified information” and a threat is defined as “an indication of impending danger or harm; or one that is regarded as a possible danger” (Cole & Ring, 2006). Insider threats are security risks that originate within a targeted organization and typically involve a current or former employee or business associate who has access to sensitive information or privileged accounts within an organization's network. These individuals tend to misuse this access. Today, there are growing concerns that insider threats pose one of the most significant risks to private and public organizations. Despite the detection tripwires in these organizations for monitoring employees’ work activities and rigorous background checks, recent reports concur that there is an increase in insider activity cases (Poll, 2015). Detection tripwires cannot always be proactive; therefore, many insider threat situations are actualized after the harm has already transpired and
sensitive or classified information has been compromised. When hiring American federal workers, organizations have maintained a “trust but verify” culture for years. However, this culture has not resolved the insider threat problem. This practice establishes routine and random auditing of privileged responsibilities. Organizations usually perform a substantial initial background investigation. Still, even after these initial and subsequent background checks, they are often entirely unaware of events occurring in their employee’s life that could make them more susceptible to being an insider threat. Trust scores may aid in identifying insider threats before they can materialize and compromise the organization. Insider threat detection effectiveness based on trust scores would likely connect to the methodology of processing employees' data through a scoring algorithm and continuously monitoring them.

According to the CERT Guide to Insider Attacks, the perceived risk of insider attacks increases, and behavioral monitoring is expected to be increased due to behavioral precursors' detection (Cappelli et al., 2012). Learning more about the differences in employees’ behavioral characteristics in the form of a social trust score may help detect insider threats. For example, an organization could increase the monitoring and observance of employees' behavior through social media when heightened intelligence of insider threats is received. If employees planning the insider threat were unaware of the various times of the heightened monitoring, they would need to be on the defensive
constantly. Utilizing social media continuous monitoring for employees may minimize their efforts, slow the insider threat down, and hopefully reveal them.

Organizations allowed employees access to sensitive information but always sought to verify their trustworthiness and monitor them. Absolute trust does not exist in the environment. Therefore, employees working around sensitive information and infrastructures will always need monitoring. The question is how and to what extent. The value of an organization is as valuable as the information they protect.

**Literature Search Strategy**

The search involved locating insider threat and trust score related literature in online libraries. The researcher used six popular libraries to search for literature: (a) ProQuest dissertation and Journal online library accessed by Marymount University, (b) the Association for Computing Machinery (ACM) online digital library, (c) the Institute of Electrical and Electronics Engineers (IEEE) online digital library, (d) Google Scholar database, (e) Sage journal database. The search strings used by the researcher were (a) “trust score” AND “Insider threat” (b) “Social Credit” AND “Insider threat” (c) “Social trust scores” AND “insider threats” (d) “Social trust score” AND espionage.

This literature review will discuss trust scores used to detect or aid in insider threats, including insider threat tripwires, and any methods of detection of trust scores in use at the time of this dissertation. This study will further advance the understanding of insider threat detection methods by analyzing ways social trust scores can aid in the early detection of insider threats. Through continuous monitoring of social media, behavioral characteristics, and technical retrieved information, many organizations will benefit from
this research and the processes this study will encompass. In a culture where trust is essential in organizations and insider threats are becoming more prevalent, it is crucial to conduct studies to help detect insider threats successfully.

Many companies and governments lean toward using trust scores to rank customers, employees, and citizens. The research will discuss how trust scores affect Human Nature, Environment, Psychological Forces, and Social Dynamics Theory. The examples included will include models for each theory and explain how an individual’s background, experiences, and group involvement influence their behavior.

**Human motivation**

In this section, the researcher discusses human aspects behind the reasoning of an insider. In Maslow’s Theory of Human Motivation, human nature is divided into five categories: physiological, safety, belonging/love, esteem, and self-actualization. The theory states that all individuals have an inherent need to belong socially, economically, and psychologically to groups (Maslow, 2017). The insider will always want or desire something that is missing in their life socially, psychologically, or economically, serving as a significant motivation for their behavior. While insider behavior is typically motivated, it can also be biological, culturally, and situationally determined. Human characteristics sometimes overshadow a specific need in which change, gratification, safety, or self-esteem are at the forefront (Maslow, 2017). Individuals feel that sometimes change is necessary for them to grow instead of them doing what is needed and required.
in an organization. This can be construed as another unusual characteristic of certain humans when their thinking is predominately overshadowed by a need that the whole philosophy of the future also tends to change (Maslow, 2017). Gratification is a more physiological need, enabling the emergence of other social goals. When the goals are chronically gratified, they cease to exist as a barrier to stopping the insider from going beyond the threshold (Maslow, 2017). A gratified insider no longer needs esteem, love, safety, etc. They might only imagine them in the metaphysical sense (Maslow, 2017). In modern times, evidence of these characteristics spills out publicly and becomes open-source data to analyze. The need for insiders to belong socially, economically, and psychologically to social media groups has opened their motivation and personality up to scrutinization by learning software and sophisticated algorithms capable of processing millions of online posts.

**Contemporary Context**

According to Nick Pickles, a senior strategist on Twitter's public policy team, insider threat detection by utilizing social media data has become an essential field of study during the expansion of collaborative websites (Bickert et al., 2018). Ninety-five percent of terrorist accounts are detected with seventy-five percent removed before they can send out Tweets. One reason is that social media is an enormous repository of information that connects people through interactions, behavior, and relationships. According to 150 Federal IT cybersecurity professionals surveyed in a federal insider threat report, organizations continue to experience insider threat incidents, despite the growing focus on insider threats and the use of detection devices (IBM/ObserveIT, 2020).
The continuous analysis and monitoring of social media data through machine learning algorithms may prevent individuals from being hired, therefore, not allowing them access to sensitive information or deterring them from acts of espionage. Similarly, the FICO score monitors credit behaviors and issues a credit score. Those individuals with low scores are prevented from receiving low-interest rates, being hired, and living in specific high-cost neighborhoods.

**Historical Background**

Trust originated as an expression of confidence between two parties in exchange for some belief that they will not be harmed or put at risk by the other party's actions or assurance no party involved will exploit their vulnerability (Sabel, 1993). During 1914 in World War I, the British and German armies gave confidence to one another to cease military operations on Christmas eve and Christmas day (Weintraub, 2002). Both sides astoundingly fired flare signals before operations started again so that men could know to get back in their trenches (Weintraub, 2002). Several researchers have alluded to trust as a multidimensional construct while discussing how expectations underlying confidence affect subsequent behavior. Barber (1983) explained an exploratory perspective of trust as a cultural and social structural phenomenon that operates in many ways (Barber, 1983). He illustrated how trust and distrust, their functional equivalent, permeate almost every aspect of our lives. In 1983, lamenting the erosion of trust without considering the levels of trust needed and desired variable conditions and valuable alternatives was
In 2004, useful options were used in a study (Pikkarainen et al., 2004) to gain customers’ trust. Using confirmatory factor analysis, the researchers developed a model indicating online banking acceptance among private banking customers in Finland. They concluded that the amount of information on internet banking positively affects consumer acceptance (Pikkarainen et al., 2004). The more details of internet banking a person has, the more likely they will be encouraged to use and trust online banking. Many American organizations today expect that individuals will be harmless, working around sensitive information, but not all do, giving rise to the "insider threat" issue (Gill & Crane, 2017).

The Chinese government has made it mandatory for all its citizens to have a national trust score by 2020 (Chen et al., 2011). Chinese officials say it is a way to influence citizens' behavior to benefit society and move their country forward. Still, some think it is just the latest step in the country’s long history of state surveillance. European countries, particularly Germany, view personal information protection as an issue of the most significant concern (Singh & Hill, 2003). Findings substantiate that consumers in Germany have extreme opinions about protecting their privacy (Singh & Hill, 2003). They believe both private sector and government organizations are obligated to safeguard the personal information of consumers and citizens (Singh & Hill, 2003). On the opposite spectrum in America, an organization typically creates rules that explain how sensitive information should be shared and to whom. These rules or detection devices have not slowed down the misuse of trusted information among employees.
Review of the Literature

Kandias et al. (2013) empirically tested a new social media analytics method, the trust scores in social media (TSM) algorithm, for measuring individual users’ trust levels in a social network. Researchers performed a correlational study and showed that social media information on Facebook or Twitter could be used to predict and deter insider threats. The proposed methodology used machine learning techniques with a dictionary-based approach to detect Greek users’ negative attitudes toward law enforcement officers (Kandias et al., 2013). Kandias et al. (2013) included 65 derogatory terms and phrases which users used to describe law enforcement and authority figures.

Alahmadi et al. (2015) focused on insider threats and proposed various deterrence and monitoring mechanisms. Alahmadi et al. (2015) used preventive monitoring and analyzed user characteristics and behaviors to inform organizations about the possibility of an insider threat breach before it happens. Psychologists have advocated that the behavior and preferences of a person can explain personality traits. This research could be leveraged to indicate the likelihood of an individual becoming an insider threat. When a user’s browsing behavior deviates over time, this research suggests the possibility of an insider threat before significant damage occurs (Alahmadi et al., 2015).

Personality has been studied extensively and is related to many aspects of job performance (Barrick & Mount, 1991). Barrick and Mount (1991) analyze user characteristics and behaviors to inform organizations about the possibility of an insider
threat breach before it happens. Barrick and Mount (1991) do not define common baseline data from a previous group of insiders convicted and found guilty in a criminal court system or employees proven not to be insiders.

Kelley et al. (2018) utilized mouse tracking in cybersecurity research, encompassing assessing risky online behavior to detect insider threats. It remained an empirical question to the extent of real-time information from tracking mouse activity and user behavior predictive of user outcomes. It demonstrated added value compared to traditional questionnaires that are self-reported. Participants in this experiment chose to "log in" or "back" out of each website based on the perceived website security. The user's mouse movement activity information collected throughout the entire process was the basis of the experiment. These participants were recruited from Amazon’s Turk using a protocol approved by Indiana University’s Internal Review Board. Participants could participate in the study if they were at least 18 years of age, spoke fluent English, and were required to use a Mozilla browser. The framework tested an insider threat hypothesis based on predictive capabilities enhanced by integrating employee and psychological data with traditional cybersecurity audit data commonly used by cybersecurity analysts. The study sought to address the work done regarding environmental and ethical issues when an organization uses employee data to predict sabotage or espionage.

In a case of the city of San Diego v. Roe, personal privacy and freedom of speech was tested against expression of personal information. The respondent described as Roe sued alleging that his First and Fourteenth Amendment rights to freedom of speech were
violated when the city of San Diego terminated his employment as a police officer, for selling police paraphernalia and sexually explicit videotapes of himself engaging in acts the city viewed as inappropriate (Rieland, 2005). A Federal District Court granted the city's motion to dismiss, but the Ninth Circuit reversed this motion, holding the officer conduct fell within the protected category of citizen commentary on matters of public concern (Rieland, 2005).

However, the U.S. Supreme Court held that the city was justified in terminating the officer whose conduct brought shame amongst the city police department and disgraced the professionalism of its officers into serious disfavor (Rieland, 2005). While Roe's speech did not relate to the workings or functioning of the police department, the speech was nonetheless clearly detrimental to the department. Further, Roe's distribution of information did not qualify as a matter of public concern and thus the city, as a public employer, was entitled to restrict the officer's speech to allow proper performance of the city’s public functions.

In the ruling, public concern was described as something that is the subject of legitimate news interest, such as negative comments about the President of the United States. Expression of information may touch on matters of public concern and should be subject to balancing of public employee and employer interests. This shows that if an employee posts a comment on social media that bring shame to an organization and
prevent it from properly performing its business duties, this employee can possibly be held liable.

Fujii et al. (2019) proposed a scoring methodology for detecting potential insider threats. The methodology constructs a model based on endpoint logs, then scores the suspicious activity based on the model. When the model detects a high score, it is classified as a potential threat. The model notifies human analysts with visualized logs to support its triage and security operations. Fuji reported on the accuracy of insider threat detection, processing time, and interpretability of suspicious activity.

Park et al. (2018) applied the social bound theory (SBT) in their research to determine if users are insider threats. This theory states that the more individuals interact with others, the less likely they are to behave abnormally, and therefore, they are unlikely to participate in an insider threat. Park et al. (2018) analyzes individuals charged or convicted of being an insider threat and compare these data points to a sample drawn from Facebook users. Park et al. (2018) strengthen the SBT methodology as the charged or convicted employees establish a baseline against which to analyze data collected from social media.

Kandias et al. (2013) proposed a machine learning model that crawled YouTube and formed a community of users with the data categorized into two segments: users with positive and negative feelings toward law enforcement. Kandias et al. (2013) focused on Greek users for their study and drew the user's information from Facebook. Kandias et al. (2013) classified data into user-related information, video-related information (number of likes or dislikes received video's license numbers, etc., and comment-related information.
The machine learning tool used text classification to identify any negative jargon directed at law enforcement or authority from a Greek username. A random unidentifiable identification number replaces the username collection for privacy reasons.

Park et al. (2018) in attempt to find the types of employees at high risk for the organization, analyzed one million tweets by sentiment analysis methodology. The tweet dataset to be analyzed was made available by a web service "Sentiment140". User tweets consisting of negative sentiments were classified by Park et al. (2018) criteria as possible malicious insiders according to the threat level. Park et al. (2018) used the open-sourced machine learning software Waikato Environment for Knowledge Analysis (WEKA) to find the possible malicious insiders. Decision Tree had the highest accuracy among supervised learning algorithms at 99.7 percent accuracy and K-Means had the highest accuracy among unsupervised learning at 95.6 percent accuracy. Park et al. (2018) experiment analyzed emotions that affect individuals’ behavior for detecting insider threats. The sentiment level and ratio of negative emotions attributed to the analysis of tweets.

Summary

In summary of this chapter, it has reviewed the background of social media being used to reveal possible insider threats and evolution of trust scores and methods of how functions within an intranet network have been used to detect insiders. This chapter has also examined the methods and effects of emojis, or reactions being used in the
evaluation of posts from social media. This chapter has also briefly examined research that used trust scores in categorizing posts like our research methodology. A group of researcher’s perspective about psychological contributed to overcome the limitations of technological approaches pertaining to incorporating reactions to the overall analysis of a post by understanding the emotions of insiders (Alahmadi et al., 2015). Their approach together with system behavior analysis substantiate the importance of using reactions in the overall evaluation when available from posts.
Chapter 3: Methodology

In this chapter, the reasoning for research approach in reducing the insider threat problem will be discussed. Data collection methodologies will be discussed pertaining to population, sampling and sampling procedures, and data collection from Facebook sites to construct a baseline trust score and sites used to analyze the main social media data. This study uses a quantitative approach for measuring each dataset and explains the research method used in detail. Next, the research explained the sample selection, the instrument and code used, and the collection of data presented. The chapter discusses a data analysis plan by explaining statistical capabilities and trustworthiness of the tool used to analyze the data. Finally, the chapter concludes with any ethical considerations which is essential in any social media research; due to the various enormous amounts of information capable of being collected about individuals online.

Rationale for the Research Approach

An experimental study of the Peace Corps Facebook page data was appropriate for this study to develop a metaphor or reference point of trust scores, since these users are held at a higher ethical standard by fellow members online (Peace Corps (US), 2006). The principal reason for selecting a Peace Corps Facebook page for constructing a baseline trust score from posts were because of the organization member’s ethics in helping a diverse spectrum people and ethical responsibility of not getting involved in public political discussions (Peace Corps (US), 2006). The posts from this Facebook page
were chosen in determining a baseline score because all their members are expected to abide by established standards of moral and ethical conduct in person and over technology (including social media). These standards are listed below:

1. Communicate courteously, professionally, and empathetically;
2. Resolve any differences through open and respectful dialogue and actions;
3. Avoid gossip, rumor, or personalization of conflicts, and the use of profanity or other offensive words or phrases;
4. Comply with policies and guidance that ensure safety and security, and protect and preserve the reputation of the Peace Corps.

All Peace Corps programs, trainings, and administrative staff and contractors are also responsible for immediately reporting any non-compliance or violation, or suspected non-compliance or violation, of the Peace Corps Policies (Peace Corps (US), 2006). If volunteers and trainees are held accountable by their fellow members on social media sites, they are not likely to write posts containing insider information, threatening comments, or unethical comments. Therefore, this page is ideal for being used to create a baseline trust score for measuring posts on other Facebook social media sites. If the trust scores from the other Facebook page posts are below the average baseline score, the posts may be seen as possibly concerning and need further analysis. This experimental study primary purpose was to develop a baseline trust score and categorize posts into categories. When these scores from posts are grouped together and analyzed using the scoring algorithm and baseline score, they will aid in determining if certain user’s posts
retrieved from CNN and Fox News Facebook pages need closer analysis for insider threat concerns.

There are over 240 million Facebook users in North America, which Fox News and CNN Facebook pages receive over 30 million of these users as visitors a day (Social Media Fact Sheet, 2022). While users may be visitors of both pages, the sheer number of users gave the study access to a vast number from the total Facebook users in North America.

**Population**

This study's proposed population is 180 million Facebook users in the United States (Social Media Fact Sheet, 2022). In 2019, an estimated sixty-seven percent of Americans used Facebook regularly and the use from the population was anticipated to increase to sixty-nine percent in 2025 (Statista, 2020). The social media platform was selected because professional employees widely utilize Facebook, and posts can be freely analyzed using tools that leverage the open Application Programming Interface (API), like Facepager.

**Sampling and Sampling Procedures**

As shown in figure 3, YouTube is the most prevalent used by eighty-one percent of American social media users. The runner-up with sixty-nine percent of social media users on its platform is Facebook, as shown in figure 3. Unlike Facebook, which relies on text and pictures, Instagram is majority a visual platform. While the sole purpose of the
platform is to allow users to share images and videos, forty percent of American social media users utilize the platform, as shown in figure 3. LinkedIn social media site whose mission is to connect the world’s professionals in an effort in making them more productive and successful is becoming similar to Facebook with the capability to share, comment, and select reactions. As shown in figure 3, this professional career social media platform is visited by twenty-eight percent of social media users. As shown in figure 3, twenty-three percent of American social media users use Twitter to send out or read tweets. A microblogging service where users can send updates by "tweets" to a network of followers from an internet enabled device. These microblogs are text-based with a limit of 140 character in length. While the Facebook platform allows a user to select a permission category (public, friends, friends with exception, only me, specific friends, custom) for sharing, the default setting for tweets is public. The social media platform TikTok which gives users an outlet to express themselves through dance, comedy, lip-syncing, as well as create videos to be shown across social media communities has twenty-one percent of the social media user community accessing it, as shown in figure 3. Lastly coming in at thirteen percent as shown in figure 3, the neighborhood social media platform Nextdoor, which is used by communities such as small towns and Homeowner Associations (HOAs). It’s used to locate lost pets, sell or donate furniture, and notify other neighbors of any suspicious activity.

The sample population studied will consist of approximately 400,000 users posts from Facebook pages. The sample population determining factor is a social media fact sheet from PEW research stating that approximately sixty-nine percent of Americans had
used Facebook (Pew Research Center, 2019), many of which work in organizations containing sensitive information.

**Figure 3**

*Percent of U.S. adults using social media (data from Pew Research Center)*

![Percentage of U.S. Adults Using Social Media](image)

*Note.* Each year within the chart represents the amount in millions which cost the companies collectively for Insider Threats. This cost has increased every year since 2017 with the year of 2020 exceeding a billion dollars.

**Data Collection (Primary Data)**

The first phase of this research will collect approximately four-thousand posts from the Peace Corps Facebook page. The Facepager tool will be used to export the data collected from posts to a csv file. Next, machine learning will be used to analyze this file, and calculate trust scores for each user’s post. When these scores are analyzed and averaged together, they will function as a baseline for determining if users are possible insider threats. In the second data collection process, Facepager will be used to collect over a half million Facebook posts from the CNN and Fox News Facebook pages. These
posts will be exported to a csv file and analyzed by using machine learning to arrive at a trust score for each post. Trust scores from these posts will be compared to a Peace Corps base line trust score, which will determine if the user’s posts are possible insider threat concerns.

**Data Collection Procedures**

In this experiment Facepager was used, because of its ability to extract data from various social media platforms. These platforms include YouTube, Facebook, Amazon, and Twitter pages, which the tool can scrape data (Saeed et al., 2022). While most social media companies have locked down and prevented the collection of personal identifiable information, the posts, comments, and tweets remain collectable by Facepager. Facepager was downloaded on the experiment’s personal computer by accessing the link located at [https://github.com/strohne/Facepager/releases](https://github.com/strohne/Facepager/releases). At the time of this research paper, Facepager Setup version 4.3.10 version was the latest version available. As shown in Figure 4, the file structure shows where the executable file had been downloaded. Clicking on the [Facepager_Setup_4_3_10.windows.exe](https://github.com/strohne/Facepager/releases) executable file allowed the Facepager program to install inside the download folder, as shown in figure 4. Once downloaded, the executable will need to be ran before this experiment can access the software program.
Figure 4

Facepager executable download

Note. The Facepager application will download in the Downloads folder by default.

Facebook have locked out Facebook accounts in the past for querying pages in a suspicious manner and labeled them as performing suspicious activity (Mancosu, 2020). While accessing data on Facebook, Facepager and other web-scraping tools are expected to follow data scraping procedures and comply with terms of service (TOS) from the platform which the data are being collected (Mancosu, 2020). If any procedures are violated, the user’s social media account that is linked by their developer’s account can be liable of being suspended or blocked on a permanent basis (Mancosu, 2020). Procedures can be something as simple as attempting to web-scrape posts above a specified limit explained in the terms of service guidelines (Mancosu, 2020).

Because of these risks, there was a new Facebook account created and associated with a Gmail account used specifically for this experiment. The email account was entered during creation of Facebook developer account as a two-factor authentication and
contact email. Subsequently, querying data on Facebook with Facepager will require the user to obtain a developer access token. The Facebook developer site used to obtain an Access token was accessed by going to https://developers.facebook.com/apps/, by selecting the Create App button on the right of the page, as shown in figure 5. This developer account will require the management of Facebook pages, Application Programming Interfaces (APIs) and retrieval of data using available restricted Facepager permissions to the social media data, so “Business” was selected as app type and the “continue” button was selected to proceed to the next step, as shown in figure 5.

Figure 5
App type selection for Facebook developer

Note. The business app type was selected for scraping data from social media

As shown on figure 6 for the Create App screen, “App Display Name”, “App Contact email” were filled out and “Yourself or your own business” was selected for the
**App Purpose.** As shown in figure 6, the App Display Name will display for the developer app to distinguish between various apps the user may have created. If there is information that need routing to the App creator, it will be sent to the app contact email, as displayed in figure 6. This experiment elected to not enter a business manager account, since this field was optional. Selecting the “Create App” will open a dialog for developer to authenticate into Facebook with a password, as shown in figure 7.

**Figure 6**

*Entering the app name and purpose*

*Note.* The “App Display Name” will be distinguished your apps from one another.
As shown in figure 7, the developer required signing into Facebook for security purposes. Once authenticated into the Facebook account and validated by any two-factor authentication, the developer app will be associated to the Facebook account currently used to log into Facebook. As shown in figure 8, the developer *App Identification number* (App ID) will be listed below the App Display Name and *Business* is listed as the type.

*Figure 7*

*Re-enter password to authenticate*

![Password Re-enter](image)

*Note.* Password will need to be re-entered to access Facepager application
This research determined the Peace Corps Facebook page would be ideal in creating a baseline trust score for classifying the posts on the other two pages as being possible insider threats. As shown in figure 9, the Peace Corps Facebook group page was accessed by going to the link located at https://www.facebook.com/peacecorps. All Peace Corps members, contractors, and administrative staff are strongly encouraged and responsible to immediately report any non-compliance or violation, or suspected non-compliance (Peace Corps (US), 2006) to the Peace Corps Country Director (CD). When CDs are made aware of any failure to comply with policies or violations of policies by volunteers or trainees, the CD is expected to take appropriate disciplinary or corrective actions, up to and including administrative separation (Peace Corps (US), 2006).
Figure 9

Peace Corp Facebook group page

Note. The Peace Corps Facebook page with the Uniform Resource Locator (URL) displayed

Every social media page has its own unique identification number, and this number usually cannot be located on the actual page, so the experiment accessed the Facebook ID finder located at https://lookup-id.com to retrieve this identification number. This is achieved by copying & pasting the Peace Corps site Uniform Resource Locator (URL) link https://www.facebook.com/peacecorps into the “Facebook ID finder” as shown in figure 10 and selecting “Lookup”. The result was the unique ID (110634980913) of the Peace Corps Facebook group page. This Facebook ID will be used to extract information from the Peace Corp Facebook page using Facepager.
Now that we have the Facebook ID of the baseline page in which we will be using to measure the other individual Facebook page posts of possibly being an insider threat concerns, Facepager needs opening in order to insert the first Facebook ID as a node. The Facepager application downloaded earlier was opened by clicking on the “windows icon” on the bottom left of the screen or typing “Facepager” in the “search window” beside the windows icon on the computer used in this research and clicking on the “Facepager app icon” underneath in the matching results. As shown in figure 11, since the experiment is using Facebook to collect data, the Facebook module tab was selected. As shown in figure 11, the “Base path” should be selected by default with our URL location and version of Facepager installed. As shown in figure 11, the data being retrieved will consist of posts, so in the resources field “/<page-id>/posts” should be selected for “Resource”. As shown in figure 11, the “parameters” should default to “<page-id>” and “<Object ID>” in the adjacent field. Maximum pages set at 10 will ensure that the query results will retrieve up to ten pages each time data is queried. The Access token obtained will need entering to query data from Facebook pages. This token can be copied
from the previous step and transferred inside the Access token field. Once the token is verified, select “Login to Facebook”, as shown on figure 11.

Figure 11

Logging into Facepager application tool

Note. Copying the Developer token App ID as shown in figure 8 into the access token text box will allow the developer to log on to Facepager for accessing data.

Next, there are two steps which needed to be accomplish before this experiment can fetch data. The first step would consist of setting up a new database for storing all the new data the Facepager tool collects for this web-scraping. Figure 12 shows the folder file structure used for saving the database. As shown in figure 12, select the “New Database” tab up on the top left of the dialog and enter in a name for the database. Make sure “DB files (*.db)” is selected for the “Save as type”. Once the “Save” button is selected, all the web-scraped data will have a database location on this experiment’s computer to be housed.
The new database was created, and the Facebook account we are logged on with is using the Facebook developer account identification number. Next, the first node will need creating to extract information based on the Facebook group page Facebook ID retrieved from a Facebook ID finder in the earlier step. As shown in figure 13, the “Add Nodes” tab was clicked, so the Peace Corp group Facebook ID could be inserted into the Add Nodes dialog box. The CNN and Fox News Facebook identification numbers will be inserted later in the experiment, by copying and pasting the identification numbers on separate lines inside the “Add Nodes” dialog box, as shown in figure 13.
Figure 13

*Add Nodes by copying in Facebook ID from Facebook Finder*

![Image of Facepager 4.3 with Add Nodes dialog box open, showing one Object ID per line with the object ID 110634980913 entered.]

*Note.* Creating a node to a Facebook page by copying the Facebook identification number one line at a time.

After selecting the "OK" button, the first seed pertaining to the Peace Corps Facebook group will be loaded in the Facepager tool, as shown in figure 14.

Figure 14

*Peace Corps Node created from Facebook ID*

![Image of a Facepager tool showing a table with columns for Object ID, Object Type, etc., with one row containing the object ID 110634980913 and the note "seed."]

*Note.* The top-level node or seed has been created so that data can be retrieved from the Facebook page.
To get the tool to retrieve specific information needed for this experiment, parameters will require inserting inside the parameter section of the tool. Therefore, the parameters needed inserting into the parameters field using presents, by selecting presets tab on the top right of Facepager, as shown in figure 15.

The parameters field section of Facepager determines information collected from social media pages. They are connections to the Facebook Application Programming Interface (API). The use of the Facepager presets, allows the insertion of parameters needed without familiarity of API connections to Facebook.

The presets tab on the ribbon will show the presets dialog box, so that existing formatted parameters can be added. Because the experiment is using Facebook, the Facebook section expanded, and “2 Get Facebook posts” was selected, as shown in figure 15. The parameters were selected and added one at a time to the parameter’s field, by selecting “Apply”, as shown in figure 15.
Figure 15

Using presets for retrieval of parameters

Note. Expand the Facebook preset to view and copy parameters into the preset parameter input box.

The parameters were able to be manually edited or removed by selecting “Edit preset” parameters, as shown in figure 15. The resulting parameters should be listed in the format, as shown in figure 16.
As shown in figure 17, once these parameters were added to collect posts, the base path and resource field was set at “https://graph.facebook.com/v3.3” and “/<page-id>/posts”, respectively. Time frame initiation was set to since “01-01-2015” until “01-06-2015” and incremented five days for every fetch of data., which can be seen in figure 17. The limit set to “50” to query posts, and maximum pages set to a “10”. This ensured a substantial number of posts from the five-day timeframe would be retrieved. Any duplication of posts retrieved will be removed with R programming during the analysis phase of this experiment. This approximate Six-and-a-half-year span resulted in 4,939 posts for the baseline Peace Corp data.
Figure 17

Parameters entered inside Facepager Application Programming Interface (API).

Note. The Limit parameter will retrieve a maximum of 50 posts for the node each time Fetch Data is selected.

Figure 18 shows posts retrieved by their posting date and displayed in the column order specified from the Facepager fields parameter setting. The query collected the messages, creation date and time of messages, and date and time messages were updated.
As shown in figure 19, on the top right side of the tool, there is JSON (JavaScript Object Notation) output table in relation to each message, which can be used in creating a template to display the data. As shown in figure 19, when this JSON was configured to include the needed fields of data, each message selected displayed the same format, but a different message, created_time, updated_time, and id. This JSON will allow the message and associated data to be displayed and exported into a comma-separated value (csv) file for analysis.
In the presets and JSON output dialog box as shown in figure 20 below, the custom table columns specify the information shown in Facepager main window. However, the presets selected previously did not retrieve an adequate amount of pertinent information needed for this experiment. For example, it did not give us the reactions. To obtain more data from posts, the expansion of fields parameter was required. The next step consisted of identifying which parameters this experiment needed to use for collecting additional data. After selecting the presets tab that allow the retrieval of column reaction parameters, the “Facebook” category was expanded and “Get reactions (v13.0)” was selected, as shown in figure 20. Once scrolling down to the column section of
the right-side of the presets window, reaction parameters can be added by selecting each parameter and selecting “Apply”, as shown in figure 20.

Figure 20

*Reactions presets added to parameters*

*Note:* These reactions from these parameters will be used in calculating a score for each post.

To go a little deeper into what each parameter does we copied them from the presets into table 1 for simplicity, which our experiment will transfer inside the fields box as shown in figure 24 to get more column information. The description column shown in Table 1 describes functionality which each parameter listed performs.
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Id</td>
<td>Id of the post</td>
</tr>
<tr>
<td>Message</td>
<td>Message attached to the id</td>
</tr>
<tr>
<td>From.name</td>
<td>Who the post was actually from</td>
</tr>
<tr>
<td>From.id</td>
<td>User id</td>
</tr>
<tr>
<td>Created.time</td>
<td>Time post created</td>
</tr>
<tr>
<td>Update.time</td>
<td>Time post updated</td>
</tr>
<tr>
<td>Type</td>
<td>For example, picture or video</td>
</tr>
<tr>
<td>Link</td>
<td>Link to the video</td>
</tr>
<tr>
<td>Picture</td>
<td>Thumb nail of the picture</td>
</tr>
<tr>
<td>full_picture</td>
<td>Full size picture</td>
</tr>
<tr>
<td>Attachments</td>
<td>If there are any attachments</td>
</tr>
<tr>
<td>Shares</td>
<td>How many shares</td>
</tr>
<tr>
<td>comments.limit(0).summary(1)</td>
<td>How many comments</td>
</tr>
<tr>
<td>reactions.limit(0).summary(1)</td>
<td>The total reactions</td>
</tr>
<tr>
<td>reactions.type(Like).limit(0).summary(1).as(like)</td>
<td>The type of specific reactions</td>
</tr>
<tr>
<td>reactions.type(Love).limit(0).summary(1).as(love)</td>
<td></td>
</tr>
<tr>
<td>reactions.type(HAHA).limit(0).summary(1).as(haha)</td>
<td></td>
</tr>
<tr>
<td>reactions.type(WOW).limit(0).summary(1).as(wow)</td>
<td></td>
</tr>
<tr>
<td>reactions.type(SAD).limit(0).summary(1).as(sad)</td>
<td></td>
</tr>
<tr>
<td>reactions.type(Angry).limit(0).summary(1).as(angry)</td>
<td></td>
</tr>
</tbody>
</table>

When parameters from Table 1 needed to be copied directly into the parameter’s field, we selected the “three elliptical dots” beside the field’s parameters as shown in figure 21 and transferred the copied parameters. These parameters allowed retrieval of social media data and storing the data in the database created earlier.
Note: Select the button with the double dots beside the current parameters on the fields row.

Since the reaction parameters were added, as shown in Figure 21, existing nodes from our peace dataset needed removing. After the parameters from figure 21 were copied and transferred into the parameter fields box, existing nodes under the Peace Corps seed were deleted by pressing “Shift button” on our keyboard and selecting the “start” and “end” of these nodes under our Peace Corp seed (110634980913), highlighting all nodes and selecting “Delete Nodes” on the top middle of Facepager as shown in figure 22, the seed will remain.
Deletion of previous nodes without reaction columns

Note. Select the first line of data and hold down the Shift and down arrow button to select the remaining lines of data to be deleted.

In the Delete Nodes dialog box as shown in figure 23, the “Yes” button was selected as confirmation to remove the existing nodes. Now that these nodes had been removed, the experiment’s next step was to produce nodes with our newly created parameters.

Confirm deletion of nodes from the data excluding reactions

Note. Before selecting Yes, be sure not delete the seed used to retrieve the data.

Additional parameters were added to the fields for the retrieval of more data, by clicking on the square ellipse with the double dots. Once the Edit value dialog opened, it
was expanded, and parameters copied and pasted from Table 1 in the dialog. As shown in figure 24, the “OK” button was selected to save the parameters, our seed highlighted, and “Fetch Data” selected for producing the new data.

Figure 24

Add additional parameters for reactions and shares

Note. After the parameters are added to the list of parameters, the fetch data button can be selected again.

Saving these parameters alone and fetching the data again will not retrieve the reactions and shares. The Custom Table Columns as shown in figure 25 will utilized parameters one key per line for specifying data shown in the main window. The existing parameters were removed from the table by highlighting the parameters in the “Custom Table Columns” and selecting “Clear Columns” as shown in figure 25.
After clearing the Custom Table Columns information in figure 25, one post was selected and highlighted to display the JSON (JavaScript Object Notation) language in the right table as shown in figure 26. This JSON output table contains all the information collected from the specific Facebook page based on field parameters. The JSON output was used to organize columns in the order they needed to display after fetching data.
In the *Custom Table Columns*, additional parameters were added one key per line to display columns in the desired order as shown in figure 27. This was accomplished by selecting the “*Key fields*” in the table containing JSON and clicking the “*Add Column*” link above the selection as shown in figure 26. Rearranging the order of columns were easily accomplished by cut and pasting them in the desired order on separate lines and selecting “*Apply Column Setup*” as shown in figure 27.

**Figure 27**

*Organize custom table columns*

![Custom Table Columns (one key per line)](image)

*Note.* Listing the columns here will display the columns in the output and ensure the data is saved into the database.

The Fetch Data query retrieved a maximum of one-hundred lines of data per page before an offcut break due to query time-out limitations and/or security features of
Facebook API (Application Programming Interface). On average there seem to be approximately two-hundred post per quarter for the Peace Corp Facebook page. For ensuring the query retrieved most message posts for this time, the search parameter fields in Facepager were started at January 1, 2015, until April 1, 2015, as shown in figure 28, and incremented three months until August 1, 2021, for every data query. Maximum pages per query was set at five. This approximate Six-and-a-half-year span resulted in slightly over four-thousand posts used in creating the baseline trust score for measuring posts on CNN and Fox News Facebook pages as possible insider threats or not.

Figure 28

CNN Facebook query data section

Note. For CNN post retrieval, the limit, since, and until fields were changed.

As shown in figure 29, on the bottom of every hundred message posts queried, an offcut for the end of a page and headers row were displayed. These rows were highlighted and deleted by pressing the Delete Node button to remove before saving to the database and exporting to a CSV file. Removing these rows, reduced chances of errors during the export to CSV and our experiment’s analysis phase.
Note. These will be entered for each query of data. They should be removed.

Now that these extra rows have been removed from the data retrieved, the existing Peace Corp data was exported to a CSV file by highlighting the seed and selecting “Export Data” button as shown in figure 31. The data under this seed were exported in bulk to a CSV file and saved into the database created in a previous step, as shown in figure 12.

As shown in figure 31, “PeaceCorpPosts” was used as the File Name, “CSV Files (*.csv)” was selected as the Files of type, the “Use a BOM” option was marked as “checked”. Unabbreviated, BOM is written as “Byte Order Mark”. It is a marker for character encoding in UTF-format. This experiment will import data in UTF-format using
R programming, so these special character encodings consisting of emojis, and emoticons will transition well. Since the order of the two bytes of special characters are arbitrary a BOM was used to mark the order. As shown in figure 31, the “Remove line breaks” option was marked as “checked” to remove possible line breaks in the posts. As shown in figure 31, each post needs a separator, so the “semi-colon” was selected as the “separator” to be compatible with R programming formatting. The data will be imported by one node at a time into different CSV files. The export mode was set at “Selected nodes (ordered like shown in the nodes view)”, as shown in figure 31. The “Save” button was selected once all needed information was completed.

**Figure 31**

*Saving and naming csv file containing posts*

![Image](image.png)

*Note.* Save the csv file inside a directory that will be accessible for the analysis portion of this experiment.

As shown in figure 32, the experiment manually saved the CSV files into a directory on the desktop named “Social Media Page Data” proceeding its exportation from Facepager. Later R Studio will be used to browse to these files for analysis.
Figure 32
Peace Corps CSV file saved to the social media page data directory

<table>
<thead>
<tr>
<th>Name</th>
<th>Date modified</th>
<th>Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>PeaceCorpsPosts</td>
<td>7/28/2021 7:46 PM</td>
<td>Microsoft Excel Comma Separated Values File</td>
<td>4.560 KB</td>
</tr>
</tbody>
</table>

Note. The data will be accessed here when needing to be analyzed.

After exporting Peace Corps Facebook baseline dataset to a comma-separated value (csv) file on our computer, the experiment entered page identification numbers for Fox News and CNN Facebook pages on separate lines inside the Add Nodes console box, so that the seeds for these pages could be added, as shown in figure 33.

Figure 33
CNN and Fox News nodes added to Facepager for creating its seeds.

Note. Type in the URL of the Facebook pages into the Facebook ID finder to retrieve the Object ID of the page, so that posts can be extracted from these sites.

Once these seeds were added, As shown in figure 34, the base path and resource field remained the same at “https://graph.facebook.com/v3.3” and “/<page-id>/posts”,


respectively. As shown in figure 34, the same parameters used to retrieve Peace Corp posts were used in retrieval of CNN and Fox News posts, so no changes in “field” parameters needed. As shown in figure 34, time frame initiation was set to since “01-01-2015” and until “01-05-2015” and incremented five days for every fetch of data. As shown in figure 34, the limit set to “500” to query posts, and maximum pages set to a “100”. This ensured a substantial number of posts from the five-day timeframe would be retrieved. Any duplication of posts retrieved will be removed with R programming during the analysis phase of this experiment.

Figure 34

Facepager settings to retrieve data from CNN and Fox News posts from social media pages

Note. The limit was increase because of expected increase posts for the five-day time span.

Post field parameters


Once all the posts were retrieved from the seeds of CNN and Fox News Facebook group pages, as shown in figure 35, the experiment ensured the resource section was
changed to “/<page-id>/comments” to retrieve the comments of these posts. As shown in figure 35, time frame initiation was set to since “01-01-2015” and until “01-05-2015” and incremented five days for every fetch of data. As shown in figure 35, the limit set to “500” to query comments, and maximum pages set to a “100”. This ensured a substantial number of comments about the posts from the five-day timeframe would be retrieved. Comments were selected and highlighted individually, then the fetch data button was selected for retrieving comments.

Figure 35

Facepager settings to retrieve data from CNN and Fox News comments from social media pages

Note. The limit was increase because of expected increase comments for the five-day time span.

Comment field parameters

On the bottom of every group of comments queried, an offcut representing the end of a page and headers row was displayed, similarly as in figure 29. These rows were highlighted and deleted by pressing the Delete Node button to remove before exporting to
a CSV file. Removing these rows, reduced the chances of errors during the export to CSV and import to the Excel file.

Like the Peace Corp dataset, the experiment exported the CNN and Fox News data to a CSV file by highlighting the seed and selecting Export Data button. The post and comment data from the CNN and Fox News seeds were imported into excel to combine this data and exported back to a CSV file to be saved into the database created in a previous step as well.

Microsoft Excel version 2016 was used to combine our saved CNN posts and comments into a single CSV file. It was used to combine posts and comments for Fox News in a single CSV file as well. As shown in figure 36, this was performed by opening Excel and selecting “Data” in the center top of the application. After the new ribbon displayed, the “From File” option was selected from the “Get Data”, and “From Text/CSV” to import data from our two csv files, which is illustrated in figure 36. As shown in figure 37, this will allow the selection of one of the files needed to import and combine in Excel. The selection of another file needing to import, and combine was selected. As shown in figure 38, the combine file was exported to a CSV file to be analyzed by exporting the data back to a CSV file by opening it in Excel and selecting “File”, “Export”, “Change File Type”, and “CSV (Comma delimited)”. This allowed the experiment to analyzed posts and comments together for our social media page. After exporting the data to a combined file, the CNN and Fox News posts combined file will be ready for analyzing.
Figure 36

*Using Excel to import CSV file for merging two files into one.*

*Note.* Performing this step allows for comments pertaining to posts to be analyzed.

Figure 37

*Selecting CSV Files to combine*

*Note.* Combining the comments with the posts CSV was be performed for both CNN and Fox News datasets.
Instrumentation

Although this research did not directly interact with humans, privacy remained a priority. The approximately half a million social media posts and comments were stored on a password authenticated computer, behind a locked door and accessed by a single user. This data did not contain any personally Identifiable Information (PII) that could be used to identify a user.

All information collected for this experiment is publicly accessible and does not require elevated permissions for accessing. A Facebook (Meta) developer’s account was granted for the experiment, which had been used to query public data. This dissertation study used the following hardware and software within this study. Aside from Internet connections, all hardware and software are considered operating within one physical location. Microsoft Windows 10 operating systems provided the base for which all research tools operated and detailed in the following section. R-Studio research Tools in Anaconda was used as an input and output tool that reads csv files uploaded to the
experiment’s computer by Facepager and provides output with both sentiment analysis and uses classification output and sentiment filtering.

Trustworthiness (Validity and Reliability)

The integrity of the experiment can be assured through Facepager’s tool used in collecting data. It specializes in collecting mainly comments and post from YouTube, Twitter, Facebook, and Amazon, but can collect attachments, pictures, emojis, and emoticons as well. The Facepager tool comes with presets, so that a researcher with little programming experience can select already constructed parameters that connects to the social media page Application Programming Interfaces (APIs). Ability to select presets could remove the error possibilities during fetching of data. Aside from helpful presets, there is a Facepager Facebook social media group which can assist with question about the tool and application settings.

Americans receive their news through a variety of ways. However social media not only gives them the opportunity to consume information, but an ever so popular way that enables them to interact and give their opinions. One of the top used social media sites is Facebook with 1.3 billion repeat users on its platform. The rationale for selecting Fox News social media page to collect data was because the sheer number of interactions it receives. In the first six months of the year in 2016, it had 120 million likes, shares, and comments on its page (Moses, 2016). These interactions well out paced the second leading Facebook page, which was NowThis at 80 million interactions within the same
six-month time frame, then next Huffington Post with 61 million interactions (Moses, 2016). The owner of Facebook itself, Mark Zuckerberg, was quoted saying Fox News “drives more interactions on its Facebook page than any other news outlet in the world. Furthermore, the rate of responses to posts are high, that shows how well information posted gets people responding with comments, emoticons, or sharing (Moses, 2016). Its interaction rate defined by all interactions divided by the total people visiting was 0.16 percent, compared to an average of 0.12 percent for fifty top social media Facebook news pages (Moses, 2016).

While Fox News is seen mainly as a conservative news outlet, CNN is viewed to many as a liberal site to access news. This makes available posts from different sectors of the American population to the experiment. In the month of May in 2022, CNN was the most monthly visited news website in the United States with 393.2 million visits per month. MSN was runner up with 347.6 million monthly visits, and next Fox News came in at 248 million visits for that same month in 2022 (Statista Search Department, 2022).

According to the literature review on "Proactive Insider Threat Detection Through Social Media: The YouTube Case," behavior analysis was performed on employees, which led to studying them under a prism of predisposition towards malevolent behavior (Kandias et al., 2013). The research examined the trait of social media and personal frustrations (Kandias et al., 2013). The most important observation is the "revenge syndrome". For example, a person developed anger towards an authority figure. The ability to use text classification to identify inappropriate user behavior is an essential tool
because it may extract conclusions for the user’s overall attitude against a supervisor, management, or government.

The social bound theory states that the more individuals interact with others, the less likely they are to behave abnormally (Kandias et al., 2013). In other words, a person who is known to interact with other ordinary people is less likely to be an insider threat. A loner may add more of a possibility of a user being an insider threat (Park et al., 2018).

Many studies regarding analyzing system behavior to detect insider threats are available, but relatively few explore an individual's emotions that affect a person’s behavior. For organizations to detect threats outside the workplace, social media analysis on behavior is necessary. According to Park et al. (2018) these behavioral findings can achieve higher detection accuracy (Park et al., 2018). The Twitter platform is widely used and can freely be analyzed using an open API. This research examined Facebook posts with an open-sourced API called NLTK Natural Language Toolkit (NLTK). The API is widely used to perform data retrieval and has multiple support documentation.

Trust scores in social media (TSM) algorithm computed the trust score for each twitter user analyzed in its experiment, which served as a proxy measure of source trustworthiness in testing the hypotheses. The score used regression predictability machine learning to predict a user's likelihood of being an insider threat, based on a decision boundary of Stochastic Gradient Descent (SGD) classifier trained with the hinge loss, equivalent to a linear SVM.
Summary

This chapter discussed the history of the trust score Implementation for Insider Threat Detection, this study's purpose, and the multiple data collection methods related to the research question. It explained the processes for detecting insider threats by retrieving posts and comments. The reasoning for selecting the sample population and factors were to protect the participants. Chapter 4 discusses the results of the trust score data analysis for all three social media groups. Chapter 5 provides interpretations of the research study.

This quantitative research will hypothesize the population I studied based on measurements of variables in the samples. This research will attempt to prove the two forms of hypotheses related to the research problem and directly affect its success.
Chapter 4: Data Analysis

Introduction

This chapter summarizes the data results received from analysis of the Peace Corps, CNN, Fox News datasets and any differences in the data. The expression used in this experiment consists of four dependent variables for building scores. These variables will be discussed and illustrated throughout this chapter. The illustration of scoring findings from three datasets determines relationship between threat categories created from ranges and perceived insider behavior on social media. These findings presented in this study appear to establish the validity of users online with low trust scores can potentially categorize them as being possible insider threats. Our categorized data findings are expected to show our Peace Corps results as keeping around the same score except for a few outliers; therefore, having an ability to serve as a baseline score for future social media results. Lastly, in summary the results in this chapter will answer three research questions:

- Could a trust score be created using machine learning techniques to reduce insider threats?
- Could a trust score be used to evaluate the trustworthiness of current and future employees?
- Could a social media page of members of the Peace Corps be used to create a model that will be useful for detecting insider threats on another social media
In efforts of answering our research questions, the study used R programming code for the analysis because of its repeatable capability (Martin, 2018). If a researcher wanted to include data from different years, they would just add the dataset, rerun the code analysis to complete data cleaning, data manipulation, data analysis, and all gets completed again at a click of a button. Another reason to use R programming is its incredible data visualization and graphics capabilities (Martin, 2018). The most important benefit of R programming is its open-source characteristics. There are thousands of people all over the world writing packages researchers can install and use that deal with specific data analytic problems (Martin, 2018).

Data Analysis

Any code used to perform data cleansing, data manipulation, data analysis, and graph creation can be accessed on GitHub:

https://github.com/michaellwilliams/Experiment_Code/blob/91ce6db724bda8261d2796ac85c39344996d8cec/InsiderThreatDetection.txt. The two-month timeframe analyzed for posts of our social media news datasets were from November 20, 2020, to January 20, 2021. The initial start of data collection fell two-weeks after the 2020 presidential election day until the inauguration of a new president, so evidently there were more data because of these major events. After accumulating posts from three social media pages, cleansing techniques were used to transform emoji icons into words, remove special characters as well as numbers from the posts. If the post was empty after removing these characters, the entire post was dropped from being included in our experiment. The data
cleansing left peace corps with 4,939 rows of data, CNN contained 49,488 rows of data, and Fox News contained 355,464 rows of data. The emoticons and emoji cons were converted into its word equivalent, as shown inside the “After Data Cleansing” table for figure 39. This allowed these converted words to be evaluated against the experiment’s algorithm used to construct trust scores.

**Figure 39**

*Data cleansing methodology for posts*

<table>
<thead>
<tr>
<th>Created Time</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-12-02 14:48:39</td>
<td>If these pardons don’t show you Trump is a criminal, your m...</td>
</tr>
<tr>
<td>2020-12-02 14:33:57</td>
<td>Hey Trump 😊 Sleepy Joe has AWAKEN 😊 😁 😁</td>
</tr>
<tr>
<td>2020-12-02 14:48:31</td>
<td>When is the second stimulus checks are ready</td>
</tr>
<tr>
<td>2020-12-02 15:03:27</td>
<td>Biden will sync our economy</td>
</tr>
<tr>
<td>2020-12-02 14:40:36</td>
<td>49 more days of love</td>
</tr>
<tr>
<td>2020-12-02 14:37:30</td>
<td>Watching the market fall will be the new norm.</td>
</tr>
<tr>
<td>2020-12-02 14:34:05</td>
<td>It’s all marginal get your berrings together!</td>
</tr>
<tr>
<td>2020-12-02 14:41:54</td>
<td>Ohh you wasted my time ... goodbye</td>
</tr>
<tr>
<td>2020-12-02 14:43:54</td>
<td>How typical surpris surpris the dji is down!</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Created Time</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-12-02 14:48:39</td>
<td>if these pardons dont show you trump is a criminal your m...</td>
</tr>
<tr>
<td>2020-12-02 14:33:57</td>
<td>hey trump sleepy joe has awaken clapping hands folded h...</td>
</tr>
<tr>
<td>2020-12-02 14:48:31</td>
<td>when is the second stimulus checks are ready</td>
</tr>
<tr>
<td>2020-12-02 15:03:27</td>
<td>biden will sync our economy</td>
</tr>
<tr>
<td>2020-12-02 14:40:36</td>
<td>more days blue heart</td>
</tr>
<tr>
<td>2020-12-02 14:37:30</td>
<td>watching the market fall will be the new norm.</td>
</tr>
<tr>
<td>2020-12-02 14:34:05</td>
<td>its all marginal get your berrings together!</td>
</tr>
<tr>
<td>2020-12-02 14:41:34</td>
<td>cnn you wasted my time goodbye</td>
</tr>
<tr>
<td>2020-12-02 14:43:54</td>
<td>how typical surpris surpris the dji is down!</td>
</tr>
</tbody>
</table>

*Note.* If the post is empty after the data cleansing, the entire post will be removed.

**Initial Score**

In the Peace Corps social media trust score dataset, it shows that most scores are within a fifty-point range, as shown in figure 40. This is expected and could be because members of the Peace Corps are upheld to a higher standard by their superiors, where they are expected not to be opinionated in a public forum, and particularly about political figures and government organizations ([Peace Corps, 2006](#)). This makes the Peace Corps dataset and trust score a great baseline to measure the other datasets trust scores. In pursuance of arriving at an initial start for scores, this experiment observed a five-year
span of time for peace corps posts used as the baseline dataset, in which the lowest trust score was 894 as illustrated in figure 40. This post was observed as being not positive nor negative by the experiment’s algorithm. As shown in figure 36, this score was rounded up to 900 for an initial start of trust scores on all posts. Figure 36 illustrates how the trust score was created and will be referred to throughout this research.

*Figure 40*

*CNN trust score post scoring example*

![Trust Score Calculation Diagram](image)

*Note.* The method used to arrive at this score will be explained next in detail.

**Average Sentiment Score**

The study took an average sentiment score of each post, which was derived from using the sentimentr package formula. This package uses ten lexicon dictionaries for the sentiment identification with eleven arguments and calculates the sentence polarity and utilizes a sentiment dictionary to tag polarized words. Sentimentr allowed sentence level posts in our experiment which contained clauses that are combined or related to receive a score from polarity. This sentimentr package is best for sentence level as compared to the Syuzhet package used to calculate the sentiment of sentences ([Shirsat et al., 2019](#)). Syuzhet package uses three methods to calculate the sentiment for National Research
Council (NRC) lexicons, Bing and Afinn (Shirsat et al., 2019). Sentimentr package calculates the polarity of each sentence with the help of more parameters like, Positive word, Negative word, Downtowners, amplifiers, deamplifiers, adversative conjunction etc. (Shirsat et al., 2019).

According to the CRAN documentation utilized by developers as a reference for R programming development, sentimentr calculates the text polarity sentiment in the English language at the sentence level. Optionally, these sentences can be aggregated by rows or into grouped variables. This experiment uses the sentiment_by function, which approximates the polarity of text by grouping variables. One of the five variables returned from the sentiment_by function is element_id, which is the original identification number of our original vector passed to the sentiment. Next is the sentence_id, which is the identification number of the posts or comments under each element_id. The word_count is summed up and added to the sentiment analysis table. The sd or standard deviation of the polarity score is included by grouping variables. The score in which this experiment used as a calculation to obtain the trust score was ave_sentiment, which took the mean of the polarity score by grouping variables.

**Insider Word Score**

This study used a list of approximately seven-hundred words from Reddit user Glorious Dawn, who found it on Attrition.org (Love, 2013). Allegedly, the overuse of these words in an email would result in the National Security Agency (NSA) to flagging
the user online as a potential threat (Love, 2013). Because the webpage containing these words were last updated in 1989, authenticity of such list should be taken with reservations, but the concept of substituting any words a researcher considers capable of being used to identify an insider threat could be viable. A half a dozen current high level political individual names and government buildings were added to this list of words and assigned sixty points. The remaining words on the list were assigned a score between twenty and fifty points, depending on perceived impact value. When search words were found in the posts as illustrated in figure 41, an insider word score was calculated based on the summation of assigned point value for these words, as shown in figure 41.

**Figure 41**

Insider word score methodology through search words

<table>
<thead>
<tr>
<th>Search Words</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>thaad</td>
<td>20</td>
</tr>
<tr>
<td>package</td>
<td>20</td>
</tr>
<tr>
<td>prime</td>
<td>20</td>
</tr>
<tr>
<td>suniac</td>
<td>20</td>
</tr>
<tr>
<td>whittier</td>
<td>50</td>
</tr>
<tr>
<td>biden</td>
<td>50</td>
</tr>
<tr>
<td>kamala</td>
<td>50</td>
</tr>
<tr>
<td>pelosi</td>
<td>50</td>
</tr>
<tr>
<td>pence</td>
<td>50</td>
</tr>
<tr>
<td>capital</td>
<td>50</td>
</tr>
</tbody>
</table>

**Post Data Cleansing**

4,939 Posts
Peace Corp

49,488 Posts
CNN

355,464 Posts
Fox News

**Insider Word Score**

<table>
<thead>
<tr>
<th>words</th>
<th>InsiderWordScore</th>
</tr>
</thead>
<tbody>
<tr>
<td>biden, fraud, trump</td>
<td>110</td>
</tr>
<tr>
<td>trump</td>
<td>40</td>
</tr>
<tr>
<td>president, trump</td>
<td>60</td>
</tr>
<tr>
<td>investigation</td>
<td>20</td>
</tr>
<tr>
<td>face</td>
<td>20</td>
</tr>
<tr>
<td>biden, face, president, trump</td>
<td>130</td>
</tr>
<tr>
<td>propaganda</td>
<td>20</td>
</tr>
<tr>
<td>trump</td>
<td>40</td>
</tr>
<tr>
<td>face</td>
<td>20</td>
</tr>
</tbody>
</table>

*Note.* The post will be searched for search words after post data cleansing to arrive at an insider word score for each post.

**Reaction Score**

This study utilizes six reactions from social media posts to impact the trust score up to a maximum of plus (+) or minus (-) two-hundred points. As shown in figure 38, an additional one point was added to the denominator equations for each reaction group
merely to prevent indeterminate readings. If there were no positive, neutral, or negative reactions, the trust score would be registered as indeterminate or “not applicable”. As shown in figure 38, the experiment took the sum of positive reactions (Like and Love) and divided the result by total reaction for the post. As shown in figure 38, for the neutral reaction, it took the difference from the two neutral reactions (Ha-ha and Wow) and divided the result by total reactions for the post. Lastly, as shown in figure 36, the experiment took the sum of negative reactions (Sad and Angry) and divided the result by total reaction for the post. As shown in figure 42, the positive and neutral reaction resultant were added, then our negative reaction result was subtracted from the remaining total, and this total was multiplied by a weighted number of 200 to arrive at a reaction score.
The six Facebook reactions users are accustomed to seeing were released in February 2016, are an extension of the single “Like” reaction seen on Facebook pages previously (Tian et al., 2017). They allow for a more nuanced expression of how users feel towards a post (Tian et al., 2017). The emotions underlying these six reactions are supposed to be frequent and universal (Tian et al., 2017). If Facebook reactions are assumed to reflect the readers’ overall sentiment towards a post, the distribution of these reactions can be used in determining the overall meaning of a post (Tian et al., 2017).

In a research on reactions and emojis for Facebook sentiment, twenty-one thousand posts, which produced fifty-seven million reactions, fifteen million shares, and eight million comments from public media pages across four countries (Tian et al., 2017), the study proposed analyzing data from four countries: UK (BBC, the Daily Mail, Daily

The analyzing of these posts sought to determine whether reactions could be generalized cross-culturally (Tian et al., 2017). While the difference across countries for reactions in the study were statistically small, “Angry” reactions were surprisingly highest from the other five reactions in France at nine percent and lowest in UK at three percent. “Love” reactions were the most frequent in US at six percent and lowest in Germany at two percent. “Haha” reactions were the highest in Germany at six percent and lowest in the UK at three percent (Tian et al., 2017).

In terms of the reaction distribution experiment performed, they found that “Like”, being the default reaction, is unsurprisingly the most frequent at an overall 78.9% (Tian et al., 2017). The frequency of the other five reactions ranks “Love” (5.5%), “Angry” (5.4%) “Sad” (4.0%), “Haha” (3.7%), “Wow” (2.5%) (Tian et al., 2017). They calculated the share to reaction ratios (amount of shares divided by number of reactions) and found them to be different for different reaction profiles. Posts were shared (16%) for “Likes/Love”, (24%) for “Haha/Wow”, and (33%) percent for “Angry/Sad” in shares (Tian et al., 2017).

These results in this experiment showed users are more likely to share a post when the reaction is something other than “Like/Love”, suggesting that stronger
emotional attitude leads to more post sharing (Tian et al., 2017). The results demonstrates that Facebook reactions and posts are a good data source for investigating indicators of user emotional attitudes (Tian et al., 2017). More conservative aligned users may visit Fox News Facebook page and more aligned liberals will visit CNN Facebook page; therefore, resulting in a skewed representation of posts and comments receiving positive reaction (Like and Love) feedback for both groups of data. However, the posts or comments receiving reactions other than Like or Love led to more sharing of content in the experiment on reactions and emojis for Facebook sentiment. The more sharing of the content, and the more comments it receives, the more possible insider threat information can be evaluated (Praet, 2022). Not to mention that both social media pages are open the public. Some Facebook pages such as National Aeronautics and Space Administration (NASA) and the National Park Service (NPS) are visited and liked at equal rates by liberal, moderates and conservatives alike (Praet, 2022). While social media information from these two pages are liked by a heterogenous audience from various political groups, posts may not trigger strong emotional attitudes which leads to comments and post sharing that are more likely to reveal insider threats. Regarding CNN and Fox News social media pages, CNN attract slightly more users who have liberal ideologies, but Fox News visitors are majority heavily “right-leaning” and conservative (Praet, 2022). Therefore, a greater number of users may agree with post through positive reactions and give the post a slightly higher point score originating from the possible two-hundred reaction points per post.

Data Results
The data in this study unveiled results for supporting the hypothesis from a trust score perspective. The hypothesis of the study needed to answer the question if a trust score could be created using machine learning techniques to reduce insider threats. The thousands of posts categorized and processed by using the Peace Corps social media baseline score, demonstrates not only could a score evaluate the trustworthiness of current and future employees, but it can also serve as a reduction of insider threats, by compartmentalizing the posts into trust score categories as shown in figure 43. In the development of our baseline trust score, we used the lowest score from the Peace Corps posts at 894 and rounded this number up to 900. This number is the threshold for our baseline score. However, to categorize the scores into five even groups for scoring purposes and because of the lowest trust score from our baseline dataset fell in the eight hundreds, posts with scores from 800 to 1200 were categorized as normal. For example, in figure 43 a post which reads “President Donald J Trump announced today his intent to nominate Alan R Swendiman to serve as deputy director of the Peace Corps” is categorized as normal with a score of 894. Eventually, these scores could be linked to employees for the employer’s review.
Figure 43

Peace Corps baseline model

<table>
<thead>
<tr>
<th>corpus</th>
<th>PeaceCorp</th>
<th>words</th>
<th>Trust_Score</th>
<th>Trust_Score_Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>president donald j trump announced today his intent to no...</td>
<td>trump</td>
<td>894.2674</td>
<td>Normal</td>
<td></td>
</tr>
<tr>
<td>on this day in congress approved legis</td>
<td>president</td>
<td>909.6828</td>
<td>Normal</td>
<td></td>
</tr>
<tr>
<td>winter is already in full force across mor</td>
<td>announce today his intent to nominate alan r swendisman to serve as deputy director of the peace corps</td>
<td>1061.4868</td>
<td>Normal</td>
<td></td>
</tr>
<tr>
<td>john mcwilliam was an education volunt</td>
<td></td>
<td>1063.7184</td>
<td>Normal</td>
<td></td>
</tr>
</tbody>
</table>

Note. One of the lowest scores in the baseline model originated from a post with a score of 894.2674 and was categorized as normal.

The central tendency descriptive table further displays the relationship between the variables, standard deviations, variance, confidence interval (CI) and standard error (SE) means, mean scores, median scores, range of scores, maximum and minimum scores of the three datasets. As shown in table 2, the minimum score for the experiment’s baseline dataset (Peace Corps – 894.27) is considerably higher than the CNN (522.52) and Fox News (492.04) minimum scores. As shown in table 2, the range of the Peace Corps was 278.28 and the maximum score was 1172. The range and the standard deviations are both smaller than the other two trust score datasets.

Table 2. Central tendency summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Peace Corp</th>
<th>CNN</th>
<th>Fox News</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum Score</td>
<td>894.27</td>
<td>522.52</td>
<td>492.04</td>
</tr>
<tr>
<td>Maximum Score</td>
<td>1172.55</td>
<td>1209.08</td>
<td>1262.12</td>
</tr>
<tr>
<td>Range of Score</td>
<td>278.28</td>
<td>686.56</td>
<td>770.08</td>
</tr>
<tr>
<td>Median Score</td>
<td>1109.19</td>
<td>909.63</td>
<td>921.33</td>
</tr>
<tr>
<td>Mean Score</td>
<td>1109.26</td>
<td>950.41</td>
<td>959.34</td>
</tr>
</tbody>
</table>
This scoring difference from CNN and Fox News illustrates an extremely low difference in ratings amongst posts in the Peace Corps dataset. As shown in table 2, the mean and median for CNN and Fox News might indicate a normal distribution, but the Peace Corps’ mean and medium are skewed, which indicates the underlying data is not normally distributed. The standard error indicates a measure of the accuracy of our mean scores. As shown in table 2, the SE mean for Fox News is a third of the Peace Corps and CNN SE mean scores.

We calculated the Peace Corps standard deviation of 15.45, as shown in figure 44, lower than the standard deviation for CNN and Fox News, 71.77 and 71.89 respectively. The standard error was calculated for all three of our trust score results by inputting a standard deviation (σ) in the numerator and the square root of total number of values in the denominator. As shown in figure 45, a standard deviation of 15.45 is placed in the equation for (σ) and the number of 610 Peace Corps trust scores was substituted for n. Once the SD was divided by the square root of 610, it resulted in a standard error (SE) of 0.6257.

<table>
<thead>
<tr>
<th></th>
<th>CNN</th>
<th>Fox News</th>
<th>Peace Corps</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE. Mean Score</td>
<td>0.63</td>
<td>0.65</td>
<td>0.216</td>
</tr>
<tr>
<td>CI. Mean .0.95 Score</td>
<td>1.23</td>
<td>1.27</td>
<td>0.424</td>
</tr>
<tr>
<td>Variation</td>
<td>238.72</td>
<td>5150.66</td>
<td>5167.990</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>15.45</td>
<td>71.77</td>
<td>71.889</td>
</tr>
</tbody>
</table>
The confidence intervals (CI) for the trust scores is a range that we would expect to contain the true population mean, and if created repeatedly would contain the true mean in 95% of the intervals. For example, the lower confidence interval score which starts at 1106.866 and upper confidence interval score which ends at 1111.654 for our Peace Corps trust scores are shown in figure 46 and figure 47. The population mean with a 95 percent accuracy for this experiment lies between these lower and upper confidence interval bounds.

**Figure 46**

*Lower Confidence Interval for Peace Corp Trust Scores*

\[ CI = \mu - Z(\frac{\sigma}{\sqrt{N}}) = 1109.26 - 1.96(\frac{15.45}{\sqrt{610}}) = 1106.866 \]
Figure 47

Upper Confidence Interval for Peace Corp Trust Scores

\[ CI = \mu + Z \left( \frac{\sigma}{\sqrt{N}} \right) = 1109.26 + 1.96 \left( \frac{15.45}{\sqrt{160}} \right) = 1111.654 \]

Variation illustrates how much each score in the group of scores differs from the group mean. The larger the number for our variation, the more the scores differ from their mean and the smaller the number, the less the scores differ. Where the variation for the Peace Corps shows 238.72, CNN and Fox News are almost identical at 5150.66 and 5167.99, respectively, as shown in table 2. This shows that the distribution of scores for Peace Corps are less dispersed from the mean. It also may reveal ranges of future scores for this group. As shown in table 3, the methodology used baselined the 610 Peace Corp posts scores as “normal”. As shown in table 3 in the CNN dataset, two posts were categorized as above normal scores, 12178 being categorized as normal, and 94 concerning. Out of 110619 Fox News posts, seven posts were categorized as above average, 110125 as above normal, 486 posts fell into the category as concerning, and one post was extremely concerning. The only post categorized as extremely concerning was from the Fox News dataset. The specific breakdown of the data is shown in Table 3.

Table 3. Frequency Results for categorized data

<table>
<thead>
<tr>
<th></th>
<th>Above Normal</th>
<th>Normal</th>
<th>Concerning</th>
<th>Extremely Concerning</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peace Corp</td>
<td>0</td>
<td>610</td>
<td>0</td>
<td>0</td>
<td>610</td>
</tr>
</tbody>
</table>
As shown in figure 48, in the Peace Corps social media trust score graph, was created from data from June 2015 to January 2021. These scores for these posts were mostly between 1050 and 1150, which were considered normal trust scores for this experiment. As shown in figure 48, the few outliers deviating from this range had a lower insider word impact score and when analyzed pose no concern. The Peace Corps graph displays the normal color-coded illustration of dots and these post’s trust scores fall inside the *normal* insider threat category. Larger dots on the graph illustrates that posts contain more insider threat words. As shown in figure 48, the graph showed that Peace Corps posts were not over opinionated in either way and did not lean toward indicating an insider threat as expected.
As shown in figure 49, the CNN social media trust score graph showed a scoring between 450 and 1350, which were categorized as normal, concerning, and above normal trust scores. As shown in figure 49, most posts fell into the normal category, but this graph showed that a noticeable number of posts from our CNN scores as concerning, and two posts were above normal. As shown in figure 49, the “insider word impact score” ranged from 100 to 400 points. As illustrated in figure 49, the higher scores illustrated the inclusion of more insider threat words within the posts. These words were assigned higher score because of the perceived threat or importance to the American society in this experiment. For example, while posts with a currently seated or living American President, “first lady”, or federal government building would accumulate fifty points for each of these words, posts containing the Terminal High Altitude Air Defense (THAAD) system word would accumulate thirty points for the “insider word impact score”. It can interpreted as the larger the dot on the graph, the larger potential threat.
As shown in figure 50, the largest Fox News dataset identified one post that was categorized as *extremely concerning*. *Extremely concerning* posts fell under 600 and *concerning* posts fell between 600 and 799. As shown in figure 50, the majority of posts, *normal* posts with scores between a score of 800 and 1199, are illustrated by purple dots. As shown in figure 50, the remaining *above normal* posts fell in between a 1200 and 1600 trust score. The graph seemed to show many concerning posts leading up to the presidential election.
Summary of Chapter 4

This chapter described the data resulting from creating the trust scores. The algorithms created relationships between posts and trust scores among the three datasets. The data used in this study was analyzed from three social media pages. The distribution of the scores plotted on graphs, categorization of scores, and the difference in the spread of minimum and maximum scores of the Peace Corps, CNN and Fox News were calculated. The Peace Corps dataset created a baseline trust scores, using machine learning techniques, that could be employed in scoring and evaluating the trustworthiness of current or future employees. If other social media platforms collect reactions from responders, the same methodology could add those platforms into the system. Any corpus of posts with reactions can be used to arrive at a trust score for each post and categorize the posts based on the baseline score created and illustrated in this study.
Chapter 5: Discussion, Conclusions, and Recommendations

The purpose of this quantitative research study was to investigate the theory that a baseline trust scores derived from Peace Corps dataset can help identify users as possible insider threats. By creating trust scores on social media, using machine learning techniques, and associating posts to users. This chapter includes a discussion of major findings as related to the literature on uses of machine learning or trust scores used in identifying insider threats from datasets. Also included in this chapter are recommendations on the handling of unrepeated or “one-offs” anomalies, where the users may have mad posts when they were unusually upset. This chapter will reiterate the discussion on reaction scores and how favorable reactions from users who normally visit social media pages and their “liking” or up-voting content can affect trust scores of these posts. The chapter will also touch on the privacy and security ramifications that may need addressing in relation to this research. The chapter concludes by discussing limitations of the study, areas in which future research can be performed, and conclude with a summary of results.

This chapter will include discussions and further research possibilities that can help answer the following research questions:

(R1): Could a trust score be created using machine learning techniques to reduce insider threats?

(R2): Could a trust score be used to evaluate the trustworthiness of current and future employees?
(R3): Could a social media page of members of the Peace Corp be used to create a model that will be useful for detecting insider threats on another Facebook site?

(R4): What are the security and privacy implications in using social data and a trust score to assess current and prospective employees?

The theory of social media user’s psychological profiles being remarkably similar who are malicious and non-malicious insider threats (Dupuis & Khadeer, 2016) can be revealed online through social media sites by creating trust scores from posts and reactions using machine learning. The psychological results of insiders are also consistent in other research observed from a personality perspective. We investigate this theory using posts on social media that can be linked to the user’s personality traits. Through assignment of scores for these posts and insider “high value” names and words, the research identified extremely concerning negative personality traits on social media and animosity towards critical infrastructure or political figures.

This research aimed at making four major contributes: (a) enable the creation of a trust score baseline with the Peace Corps dataset. (b) use this baseline trust score to evaluate the trustworthiness of users posts from other social media sites. (c) use our baseline scores as an upper and lower limit model; including the capability to categorize current and future posts. (d) investigate the security and privacy implications in using social data and a trust score to evaluate current and prospective employees.
Much of the scoring factors related primarily to the posts and comments, but reactions added a maximum of a negative or positive 200 points, as shown in figure 42. This reaction score can be weighted due to the unbalanced number of “right leaning” or conservative users visiting the Fox News social media pages, compared to CNN which has an almost equal number of users with different ideologies (Praet, 2022).

In figure 51, there is a side-by-side comparison of posts from Fox News and the Peace Corps with their calculated trust score from the algorithm. This illustration validates the ability to create a trust score, addressing research questions R1. This study created a trust score with machine learning techniques that could be used to explore insider threats through evaluating current and future employees through their social media posts. If, for example, an employee shows a negative opinion about the authenticity of voting machine software and espouses the desire to sabotaging the system, the experiment’s trust score algorithm could highlight their post and raise concern, either while an employee or during the hiring process.

Two example posts are shown in figure 51. Comparing the posts illustrates the different in tone and content between a trust score of 492 and a score of 1137. While not illegal to posts derogatory comments about political figures, e.g. former president, vice president, or make allegations about the security of the voting systems being used to decide elections, certain organizations may believe having knowledge of these sentiments might protect an organization from an insider threat.
Interpretation of Findings

While scores from posts and comments varied in this experiment, they were derived from the four common factors: (a) an initial score of 900 (b) the average sentiment (c) the accumulated insider word score, and (d) the total reaction score. These components of the score allowed for the categorization of posts into extremely positive, positive, normal, concerning, or extremely concerning. In the next section, we will compare this research with previous studies to explore insider threats.

Proactive insider threat detection through social media: The YouTube case

Kandias et al. (2013) conducted a study using machine learning techniques with a dictionary approach of Greek words. It was shown to detect users on social media holding negative attitudes toward law enforcement officers (Kandias et al., 2013). The methodology proposed employed machine learning techniques with a dictionary-based approach to detect Greek users holding negative attitudes toward law enforcement officers. The study used sixty-five derogatory terms and phrases to query users who
described law enforcement and authority figures negatively (Kandias et al., 2013). While the study was performed for the country of Greece, it solely included posting in the Greek language from social media (Kandias et al., 2013). The indigenous population speaks a variety of Albanian (Arvanitika) in Greece (Simos et al., 2014). Including only Greek language postings from users in the community may be looked upon as a weakness in the study (SIMOS et al., 2014). While the study analyzed 2,043,362 comments with machine learning techniques, the study failed to identify a baseline dataset that could be used in categorizing severities of Greek users’ negative attitudes against law enforcement officers.

In contrast to Kandias et al. (2013), a baseline dataset was used to identify possible insider threats and served as an initiator to categorizing posts. This research used over six-hundred words to evaluate threats. This study also shows situations in which “one-off” or a user who is having a “bad day” and makes a derogatory comment on social media can be handled. These situations can be handled by enabling the trust scoring methodology system to evaluate the frequency of such posts in comparison to the ones that are positive in nature from users. This can be a mechanism of identifying users unusually making a bad comment because they are upset compared to an actual insider threat.

**Using Internet Activity Profiling for Insider-threat Detection**

Alahmadi et al. (2015) used preventive monitoring, analyzed users’ characteristics and behaviors in the confines of an employer’s computer network to inform organizations
about the possibility of an insider threat breach proactively. Alahmadi et al. (2015) based their work on psychologist’s research that had shown a person’s behavior and preferences can explain personality traits. These traits could indicate the likelihood of an employee being an insider threat. According to their study, detecting the deviation of user’s browsing behavior allows the possibility to identify potential insider-threats before considerable damage is done (Alahmadi et al., 2015). While this browsing deviation over time at work can be related to many aspects of job performance (Barrick & Mount, 1991), it does not necessarily consider the employees browsing history away from their place of employment. This research analyzes user characteristics and behaviors to inform organizations about the possibility of an insider threat breach before it happens (Alahmadi et al., 2015). If the employee knows that their employer is monitoring their browser history, they may not visit certain sites all together. The research does not define a common baseline dataset from a previous group of insiders who were convicted and found guilty in a criminal court system or employees proven not to be insiders.

In contrast to the study above which employees are analyzed using characteristics and behaviors in the confines of an employer’s computer network, our study evaluates users on social media as being possible insider threats whether they are on the employer’s internet or on their personal internet. The detection of insiders is not limited to current employees but can be implemented and incorporated into an assessment of future employees. Where this research has not evaluated the social media posting of convicted
employees and this very well may be a future research topic, it has identified and created a baseline trust score using the Peace Corps group for other social media posts. This research can be expanded to include many different social media sites with millions of users, not just employees browsing on an internal internet network.

**Statistical Models for Predicting Threat Detection from Human Behavior**

Kelley et al. (2018) sought to address the legality and ethical standards concerning organizations during the collection and analysis of data to predict sabotage or espionage. Kelley et al. (2018) monitored computer mouse tracking to assess potentially risky behavior and insider threats. It remains an open question as to the utility of real-time data in tracking mouse movements for predictive outcomes of insider threats. However, Kelley et al. (2018) demonstrated added value by requiring members of the University to perform movements on a computer compared to traditional questionnaires given out to participants in self-reported experiments. The user’s mouse movement activity information collected on each site the participant chooses to “log-in” or “back” out based on security implemented was the basis of the experiment (Kelley et al., 2018). The participants were recruited from MTurk, a crowdsourcing marketplace of volunteers that makes it easier for individuals and businesses looking to outsource jobs and tasks that can be perform virtually. Following protocols approved by the Indiana University’s Institutional Review Board, the research reviewed 214 participants completion of jobs that required movement between different sites using a computer mouse (Kelley et al., 2018). Proceeding the data analysis phase, 41 participants were excluded from the experiment for not initiating the task (Kelley et al., 2018). An additional 50 participants
were excluded from analysis for only logging certain websites, but not completing an adequate number of tasks, resulting in 123 participants (Kelley et al., 2018). The framework sought to identify an insider threat, based on predictive capabilities enhanced by integrating employee data, psychological data with traditional cybersecurity audit data commonly used by cybersecurity analysts (Kelley et al., 2018).

Lam (2016) described the collection of data to detect insider threats as running the risk of infraction for discrimination, infringement on personal privacy and freedom, or hindrance of employees’ activities protected by United States law. The United States Supreme Court confirmed that adverse employment actions against an employee for off-duty and off-premises comments (Rieland, 2005), such as a post or video on social media, was justified in terminating an employee, Roe. Roe’s conduct brought the mission of his department and the professionalism of other employees into serious disrepute (Rieland, 2005). While Roe's speech did not relate to the workings or functioning of his job, the published public information was nevertheless clearly detrimental to the department (Rieland, 2005).

In contrast to the study above which used participants from MTurk to follow a protocol that required the movement between different sites using a computer mouse, this study fills in the research gap by collecting actual posts from thousands of different users on social media. While the mentioned experiment above recorded data from participants utilizing a computer mouse for a limited amount of time, this
experiment collected a variety of users post data twenty-four hours and seven days a week for multiple months. There was a baseline scoring model put in place before the analysis of the data in the CNN and Fox News datasets.

**Scoring Method for Detecting Potential Insider Threat based on Suspicious User Behavior using Endpoint Logs**

Fujii et al. (2019) is most like our research in its scoring the suspicious activity based on a model. The methodology constructs a model based on endpoint logs, then scores the suspicious activity based on this model (Fujii et al., 2019). When the model detects a high score, it classifies the suspicious activity as a potential threat on the network system. The model notifies human analysts by alerting them with visual logs to support triage and security operation (Fujii et al., 2019).

Similar to our research, the low scores are not necessarily always the result of insider threats. The categorizations of threats from the network could further be divided into severity groups, like the five groups consisting of extremely positive, above normal, normal, concerning, and extremely concerning in our research. In contrast to the research mentioned previously, our research has an ability to proactively prevent vulnerabilities through insider threats. In monitoring social media our research aims at discovering the insider threat vulnerability before damage is done, coupled with the ability to notify analyst of a possible insider threat by the scoring model.
Implications for Research

Against external threats, organizations can enable proven and established safeguards, such as antivirus software, firewalls, and surveillance cameras. There are very few safeguards against an insider threat inside an organization who can make efforts to illegally remove data and pass it to an adversary or even delete and sabotage systems. Many insider threats already have authorized privileged accounts that have access to services, servers, and databases. These assets can be bridged by multiple networks throughout an organization with some of the same administrative passwords. With the proliferation of “working remote” from the employee residence, these threats can come from anywhere. The insider can be working from inside or outside the United States for an organization. Therefore, employees need to be vetted constantly from a place where millions of employees visit, and that place is the internet.

The results of this study proved that insider threats can be identified through social media using machine learning and trust scores; therefore, adding another safeguard for detecting insider threats before and during the employment process. Detection can be done outside or inside the workplace when an employee makes a post to a social media page. These social media sites like Facebook, Instagram, and Twitter allow family, friends, and even co-workers to share the latest personal news and opinions on a variety of topics. In eighty-five percent of the incidents caused by insider threats, someone other than the insider knew of intentions, plans, or activities leading to the incident (Cole &
Twenty-two percent of co-workers and twenty-two percent of friends and family witnessed or knew about a pattern of activities that would identify them as an insider threat (Cole & Ring, 2006). Patterns that could be identified by analyzing social media pages consisting of post and comments from the insider and individuals involved in having full or partial knowledge of activities.

No matter how much a person may perceive themselves as getting away with a crime, majority of the time there are always going to be others who know what they are doing. We witness this every time after someone gets caught committing a crime and the authorities interview their friends. These friends or family members would say something like, “I knew he was up to no good because they made comments that were suspicious; this doesn’t surprise me the least”. The question most people ask themselves is: why didn’t they alert authorities or say anything? Maybe friends and family had pieces of the information and not enough they thought was required to report them, forgot to take time for reporting, or felt obligated not to report. In that case, what if it was an application which collected posts and comments from the possible insider and family and friends on social media sites, analyzed these conversations, gave these conversational post trust scores broken down into groups, and alerted authorities when a score from post met a threshold? These are all possibilities from the research of this dissertation. While it does not remove the need for co-workers, friends, and family in reporting insider threats, this research may make it more difficult for insider threats to avoid detection in the future.

One noticeable difference in the results of this study was that it categorized posts into five categories (extremely positive, positive, normal, concerning, extremely
concerning) by assigning scores similar to an American FICO credit score. This allows for other factors from society like an actual credit score, criminal record, or overseas travel, etc., that can be used to additionally evaluate an individual over all trust.

**Limitations and Recommendations for Future Research**

In the previous chapter we covered the results of our study identifying insider threats. This research identifies ways our methodology can assist in revealing insiders, but the only way to absolutely know for sure is if they are caught and found guilty. The machine learning algorithm incorporated into the research allows searching and identifying words in posts or comments. One recommendation for future research would be to include a social media page that has majority “left-leaning” and liberal visitors, like the New York Times or National Public Radio (Praet, 2022). This should counterbalance reaction feedback from the majority of users who are conservatives, visiting the Fox News social media page. While the research baseline dataset created from Peace Corps posts is used to measure and categorize posts for being possible insider threats, it’s important to include a social media page on the other side of the spectrum visited by most liberal users that will have a slightly greater impact on reaction scores for posts.

While this research collected approximately four-hundred thousand posts and comments to analyze, the results of this study were limited by users’ names not being able to be retrieved by the study’s data collection tool. Facebook currently prevents tools
from massively collecting personal identifiable information like names and locations. Knowing the location of the users who are made the posts would have aided our research in categorizing the posts origination by state, region, or country. Because personal identifiable information will be absent from this study, the Internal Review Board (IRB) was not needed to review our collection of data methods.

Another limitation of the research was the sixteen gigabytes of drive space the computer for our experiment contained. During the experiment of analyzing the datasets, a memory allocation needed to be increased from approximately sixteen thousand to almost one-hundred thousand. This increase of resources using R programming allowed the larger Fox News dataset to finish processing in two-hours. If the computer had increased resources, an increased number of posts could have been analyzed.

The R code used in this research is currently on GitHub and a future effort is the process of developing a package to allow researchers to leverage our methodology. Further work refining the workflow would simply the ability to create trust scores from a variety of social media. With the evolving nature of social media as well as trust and privacy, additional research on the desirability and social acceptability of trust scores also offer potential for future research.

**Recommendations**

This study goal was to develop a trust score for evaluating social media posts as being a possible insider threat. The Peace Corps social media posts were used to arrive at a baseline trust score for this evaluation. While the techniques used in this research can
be useful in determining users’ trustworthiness who make such posts, a great deal of care and responsibility should be taken when publishing this information to the public. If this retrieved information isn’t properly managed and protected, it could affect the possible associations or employment in an individual’s future.

That being discussed, there are times when a social media user may be unusually upset from a long stressful workday or the normal family difficulties and pose no threat to society. This could unwarrantedly cause an individual to be judged for a difficult day and may be the cause of personal situations which occasionally occur. Therefore, additional research to incorporate the number of times a user makes concerning social media posts and the level of severity should be taken into consideration.

**Conclusion**

Many American modern workplaces exhibit a secure work environment with network detection mechanisms, physical security like badge readers and cameras, although many insiders’ methods to confiscate massive amounts of data have evolved. One of the covert ways to remove media from an organization is through removable media. The same size removable media stick that could store a maximum of 128 MB (Megabytes) a few years ago, now can collect 5 GB (Gigabytes) or forty times as much data. Those organization who depends on wireless capability for their employees, the possibility of data collection may be worst. An insider can buy a wireless access point
(WAP) for fifty American dollars and in five minutes plugs Category 5 (cat5) cable from their computer into the WAP belonging to the organization, take another cat5 cable and plug it into the purchased WAP and the company now have a rouge wireless access point which can steal sensitive data from employees’ computers. In any case, companies need to be relentless to prevent intellectual property that includes intangible creations by organizations.

While there is an appearance of stringent security inside most organizations, insider threats are more difficult to prevent than the external threats that organizations face. Most security efforts have been focused on the external threats. These threats are easier to see, easier to stop, and easier to control. Unlike the ability to receive notifications about abnormalities on an organization network for external threats, there may be no alert notifications of an insider threat. A response team for the organization may not have an idea that an insider threat has been going on. Insider threats are like “fires”. The longer the insider remains undiscovered, the more damage it does.

Many current solutions do not scale to prevent insiders from confiscating information illegally. Most solutions deployed internally on organization’s systems are meant to stop external attacks from the outside. Solutions which consist of Firewalls, intrusion detection systems (IDS), and intrusion prevention systems (IPS) are meant to protect an attack vector that is already known as a weakness. Firewalls are meant to block threats entering through certain ports, which stops an outside attacker but does little for an insider. For this reason, the insider who is given all the details, credentials needed, success to attain an organization’s valuable information is almost guaranteed. If an
attacker maliciously compromises a network by attempting to get through a security system, they have a greater chance in being detected. However, an insider usually already has access to get through these security systems and have a less of a chance in getting caught.

Despite insiders being able to authenticate to system and having permissions to massive amounts of data without triggering an alert from a detection mechanism, this research has proven the ability to identify possible insider threats by using trust scores. With over eighty-one percent of the population posting on social media (Auxier & Anderson, 2021), the ability to reveal an insider from anywhere and at any time is possible. In fact, the results of this study suggested that any keyword can be incorporated into a list of words for proactively identifying the plans of an insider to commit an insider threat. This detection of insider threat methodology using social media and trust scores can not only reveal insiders plans but deter them from ever attempting to engage in these acts and can be used by human resource managers in vetting employees prior to hiring. Considering privacy regulations and laws, hopefully workplaces will use this study as a template for future implementations and variations of this research to reduce insider threats in organizations.
References


Barber, B (1983), *The Logic and Limits of Trust.* New Brunswick, NJ: Rutgers University Press. [Cross Reference]


Cappelli, D. M., Moore, A. P., & Trzeciak, R. F. (2012). *The CERT guide to insider threats: how to prevent, detect, and respond to information technology crimes (Theft, Sabotage, Fraud).* Addison-Wesley. [Cross Reference]


Curley, C. (2018). Many countries don’t use credit scores like the US here’s how they determine your worth. [Cross Reference]


Dupuis, M., & Khadeer, S. (2016, September). Curiosity killed the organization: A psychological comparison between malicious and non-malicious insiders and the insider threat. *In Proceedings of the 5th Annual Conference on Research in Information Technology (pp. 35-40)*. [Cross Reference]


Luhrmann, Niklas (1979), *Trust and Power*, New York: John Wiley. [Cross Reference]


Mancosu, M., & Vegetti, F. (2020). What you can scrape and what is right to scrape: A proposal for a tool to collect public Facebook data. *Social Media+ Society, 6*(3), 2056305120940703. [Cross Reference]


Moses, L. (2016, June 27). How Fox News became the most-engaged news site on Facebook. [Cross Reference]


Ragnell, F. (2020). Technological Salvation or Orwellian Panopticon?: A Case Study on Social Labelling, Governance, and Social Control in China’s Social Credit System. [Cross Reference]


Statista (2020). Most popular Facebook fan pages as of June 2020, based on number of fans. Statista. [Cross Reference]

Statista Search Department (2022, June 30) Leading news websites in the U.S. 2022, by monthly visits. [Infographic]. [Cross Reference]

Stephenson, C. (2020, Aug 31). Fact check: The U.S. is not the only country to have a credit score system. *USA Today*. [Cross Reference]

Thao, T., & Tsanthaiwo, S. (2017). Apple Inc. In the Years to Come. SSRN. [Cross Reference]
