THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF INFORMATION SCIENCES AND TECHNOLOGY

Analysis of ISIS Twitter Media Content

MICHAEL JAMES FANELLI
SPRING 2022

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Data Sciences
with honors in Data Sciences

Reviewed and approved* by the following:

Anna Squicciarini
Associate Professor of Information Sciences and Technology
Thesis Supervisor

John Yen
Professor in Charge of Data Sciences in IST
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

This thesis uses data from a newly collected source that is referenced here as the ISIS Twitter dataset and a Ground Truth web sample scraped from Google to create a learning model that can accurately identify the subject of a ISIS related photo (being “Military”, “News”, or “Religion”). To begin, this thesis will introduce motivation for this research was and provide a general overview of this thesis. Then it will address the datasets used as they are not in the public domain. Next previous research in the space of radical group media analysis will be discussed. An overview of the methodology for training and testing, as well as explanations of key concepts will be addressed in subsequent section. Results for the models used will be displayed and demonstrate that there is a capability to properly label images. Furthermore, conclusions from these results will be drawn. After this, limitations, and improvements for the research conducted will be addressed. Finally, future work that can be done with this thesis and a conclusion on the research will de discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

To begin, my thanks must go to Professor Squicciarini who provided me the opportunity to help with the research she has been doing in this field. This work has not only made me a better data scientist, but more importantly has made me a better scholar. Through the scholarly work I've done with Professor Squicciarini I've learned how to guide myself through the troubles and tribulations that arise with research, how to conduct research properly, and most importantly enjoy what I am researching. Another person I would like to acknowledge is Younes Karimi. Younes is a PHD student who has worked extensively on the dataset I used for testing and has provided me with some insight on how to improve my research. Younes assisted greatly in giving me a better understanding of the data and helping me to label and group the photos.

Chapter 1

Introduction

Unfortunately, many people live in the world where terrorism is a prominent issue and with the advent of social media the ability to spread messages promoting terrorist ideology and spread fear has become much easier. Social media provides for the easy dissemination of information to groups of all sizes. One of these groups that was quite active on social media when they were at their peak was ISIS (Islamic State of Iraq and Syria). ISIS, in short, is a radical Islamic group that seeks to destroy all those who oppose their interpretation of the Quran and create a new Sunni Islamic state. Understanding the type of content that gets posted by groups like ISIS may provide more insight into how terrorist groups spread information, what they care about, and may even help social media groups further identify and remove dangerous content.

The goal of this research is to provide a better understanding of the type of content that ISIS posts on social media (specifically Twitter) and ultimately create a model that can be fed in a set of photos to understand the subject of the content being posted. We hope to develop models able to produce two key findings. The first is to provide a method that can help to identify what different types of content are being posted by ISIS accounts so we can better understand the mindset of supporters of ISIS. This may lead to a better understanding of the ISIS group as a whole and also provide insight into how to research other terrorist groups. The second item that can hopefully be produced from this model is method for an automated labeling of images to create the studying of these images easier. Just in this dataset alone that is being analyzed there is over 300,000 pieces of content that are videos, gifs, or photos. For the purpose of this research the focus was on photos.

This paper is organized into a several sections and is ordered as follows. First it begins with an introduction to the dataset and the goal of the project. Then it moves into a literature review of what other groups have done to analyze terrorist social media. This paper will then move into the methodology and the results of the model and how well it can predict the subject of the photos being fed into it Finally, I will go over what could have been improved in my research and the Conclusion of the research.

Dataset Introduction

One of the key aspects that will be discussed and mentioned throughout this paper is the datasets that have been used throughout this research. The first will be referred to as the ISIS Twitter dataset. This dataset consists of tweets from 2014 to 2015 and the tweets are from accounts that are either from known ISIS accounts or ISIS-related accounts (). Of this dataset, I specifically looked at the tweets that had some form of media attached to them. The types of media that could be attached to these tweets were either gifs, videos, or photos. For the purpose of this research, photos were the main focus, however, a breakdown of the Type of Content and the Number of Pieces of Content can be seen in Table 1. There is a lot more information about this media was provided that contains more information about the different media such as user and a link to the tweet the media is attached to, however, as previously stated, the focus will be on solely the content devoid of the tweet and any other information. It should also be noted that some of the content and the tweets are no longer available for viewing as there have been changes to the Twitter terms of services, thus some of this content has been banned.

Table 1 ISIS Dataset Breakdown

Type of Content	Number of Pieces of Content
GIF	390
Photo	367,524
Video	5,814

Figure 1 Example ISIS Dataset Images

The second dataset that will be introduced in this research is the Ground Truth Dataset. Part of this research, which will be discussed in more detail in this paper, involved the generation of different groups to label the subject of these ISIS Twitter photos. The labels that were used were “Military”, “Religion”, and “News”. To help create a model that could label the photos, the Ground Truth dataset was created to train the model on photos that could be found on Google Images using key words that represented the labels. For example, a lot of the images that would be labeled “Religious” had images of Mosques. So, using the Selenium library provided in Python, an instance of Google Images can be created, and the key word “Mosque” can be input into the search bar. Once the keyword is entered into Google Images, the Selenium packaged will parse through Then a specified number of photos can be downloaded that would appear when a user looks up the word “Mosque”. This allowed for the quick and easy downloading of images that fit the labels that were generated. Table 2 shows the keywords used to generate images for the respective tags.

Table 2 Selenium Key Words

Label	**Keywords Used**
Military	*Military, Military vehicles*
Religion	*Mosque, Prophet Mohammed, Quran, Quran passages, inspirational Quran passages*
News	*News, News Reporter*

As previously stated, the Ground Truth dataset was avenue that was explored but proved to not be a viable option. A more detailed discussion of this will be addressed in the Areas of Improvement section of this paper. However, though the process of using Selenium to collect image data may have potential for other projects or may have simply been misused in this research

Chapter 2

Literature Review

The study conducted in this first piece of literature analyzed the number of Jihadist accounts that are considered resurgent accounts. A resurgent account, simply put, is an account that has been deactivated, but that user creates another account that is identical in nature. The researchers believed that there has not been enough research conducted in this area and that the number of estimated deactivated Jihadist accounts isn't quite accurate. As the research team puts it "there are no methods to identify resurgent accounts amongst this volume of data, or control for the biases that they cause" (2016), these biases being whether the all the accounts suspended were unique accounts. Because this article emphasizes trying to understand how these resurgent accounts form and what happens after they are deleted, this research would best be described as a form of analytical research. The researchers snowball sampled twitter accounts for 77 days, taking 1,000 a day, looking for accounts that were followed by 10% or more of the sampled users and had less than 1,000 followers that were not shared by other accounts (Wright, et al,. 2016). To determine if an account was a resurgent account, they compared the wording in the Twitter bio and if it shared 30% similar words it was then checked by a human to see if the accounts are indeed a resurgence of a past account. The researchers concluded that there are about 20-30% duplicate in samples and have this novel idea of these resurgent accounts can help other researchers be more aware of potential biases in their datasets.

The research conducted in this article is a little more general in its scope than the previous article I addressed. This piece of literature attempted to analyze just how successful the ISIS social media strategy is working. There were three main areas of focus for the research in this article. The first is on the measuring and evaluation of known, suspended ISIS accounts.

Alexandra Siegel, and Joshua Tucker looked at the tweets of over 16,000 accounts and identified the language and hashtags that these accounts most often use. A lot of them are words of support or to promote trending topics with hashtags like #Caliphate_News (2018). The second aspect they looked at is what is the societal view of ISIS. Researchers analyzed over 7 million tweets, again, creating an appendix of words that denote either support or oppose ISIS(Siegel, Tucker, 2018). In addition to these keywords, they also looked to see where coming from across the globe. Finally, they analyzed the type of content that is posted by both pro-ISIS and anti-ISIS groups and people on YouTube. Utilizing their appendices of key words for both support and opposition of ISIS they tried to identify content and see what the majority sentiment is on YouTube. Again, as with most, if not all articles I will address, this article uses analytical research.

This journal seeks to understand the antecedents of support for ISIS on Twitter. To elaborate, a group of researchers looked at Twitter profiles that were considered pro-ISIS and anti-ISIS and then tried to look at the prior activity accounts to try and then predict whether an account would be in support of ISIS or not. At its core, the question of the research is, how can we predict a future ISIS supporter? Since there is more of a focus on trying to understand and analyze these accounts and what can be used to predict their actions. It is safe to assume that this is analytical research. Similar to other research in this area, the researchers used the Twitter Rest API to pull around 180k tweets from an 19-day period from Arabic speaking users (Magdy, et al., 2016). They then asked an Arabic speaker to classify 1,000 of those tweets as opposing, supporting, or have nothing to do with ISIS. From this the researchers deduced that people who refer to ISIS as the Islamic State are pro-ISIS and those who refer to it as ISIS are anti-ISIS, and they used this scheme to classify more tweets. Following these classifications, they pulled tweets from suspected supporters and opposers and analyzed the tweets from these users prior to the rise in ISIS.

The results were that the classifier they created to deduce accounts that would become ISIS supporters was about 87% accurate (Magdy, et al., 2016). It appeared that many supporters were frustrated with what occurred at the Arab Springs.

The following research article that I will address discusses the sentiment of people as it regards to ISIS. To elaborate the research analyzed the text data gathered from the Twitter API and analyzed how people from eight different countries feel about ISIS. The researcher was looking at two key questions. The researcher wanted to understand how people's views of ISIS were different based on whether they were in the East or West. Then after understanding this, the researcher investigated "if there is a significant difference between the ratio of the negative and positive words of the tweets from the users from the two sides of the world" (Mansour, 2018). To conduct their analysis the researcher used a quite novice, all be it effective approach. They collected tweets from 4 of the biggest countries in the West, and 4 of the biggest countries in the East and using some Natural Language processing techniques to analyze the key words of the tweets. Then the researcher used a lexicon of words that have been determined to be positive and negative in sentiment to classify tweets and then performed TD-IDF to further analyze the key words that appeared (Mansour, 2018). What was concluded from this study was that there was in fact not much difference in the terminology people used to express their views of ISIS throughout the eight countries. Nor, was there really that much difference in the proportion of people who were pro-ISIS and anti-ISIS between the East and West.

The final piece of literature that I looked at to help me understand the topic I will be pursuing looks at what they call, an ISIS OEC. An OEC is an Online Extremis Community, which is self-explanatory. Using a complex network graph, this research group attempted "to gain insight into this online social network of unaffiliated sympathizers, propagandists, fighters and recruiters, and how these users interact…" (Benigni, et al., 2017). To do this the researchers used a complicated Iterative Vertex Clustering and Classification method. In layman's terms, the researchers first identified some clearly pro-ISIS accounts. Accounts that may simply belong to supporters, or fighters, anyone who shows clear support through their posts. The researchers then used this IVCC method to attempt to try and connect other accounts that the user follows or mentions, or by the hashtags that they use. Using this, they can then create a social netowrk graph to identify a OEC and provide the insight they looked to provide with their research. Based off this graph, the researchers managed to produce two important conclusions. The first is that they simply even made this IVCC model. This model managed to outperform two already existing models that are being used to try and model and predict these OECs. The second conclusion is that they managed to provide three key insights into how these groups can be studied. The first is that using the multimode and multiplex design of Twitter, their algorithm was able to perform much better (Benigni, et al., 2017).. Secondly, by identifying these OECs, they could more easily analyze the group. Finally, the IVCC had such great results, that it could be applied potentially at large and be used for more informed decision making.

Chapter 3

Methodology

Below outlines the entire process that was undergone through the timeline of this research. The research first began with a manual analysis of the photos themselves to get insight into the potential groups (labels) that the different photos could fall into. There were many different attempts to create a model and those approaches are detail in the Initial Approach section. In the Final Approach section, the final methodology that was used for this research is detailed.

Data Labeling

To begin understanding what type of material was being presented in the photos that were posted on Twitter, a manual review of the photos was performed. Both a spreadsheet and a folder contained all of the media that had been uploaded by the accounts that contained ether gifs, videos, or photos. With photos being the focus of this research, photos will be the only aspect discussed in the labeling section. Initially, the photos were looked at via the spreadsheet. The spreadsheet contained a lot of different information about the tweets, but the main thing it possessed was a link to the tweet itself. Viewing the tweet as well as the photo could be used to help provide more context for the image. The first round of labeling began by clicking through 493 links that were said to contain photos. Unfortunately, of these 493 photos, only 59 of the photos actually existed. This was quite common with all types of media as the change in guidelines for different social media sites (specifically Twitter) has caused a lot of content to be suspended or unavailable. However, from these photos I was able to generalize the subject matter of those photos into three categories. The three categories are Military, News, and Religion. The Military photos are ones that are considered images that contain images of military action, people in uniform, weapons, or anything that may be associated with violence The News label is used for what appears to be images from news channels, news reporters, and newspapers. Regarding Religion, these images were labeled if they contain such imagery as prayers, Quran passages, mosques, and anything that may have made reference to the Islamic religion. Obviously, not every photo could be encapsulated in these groups. However, for two reasons, photos like this were avoided. One reason was these photos were in the minority. Most of these photos could fit within the confines of these levels, though to confess it was not possible to do a full analysis of all 300,000 plus photos to determine groups that could capture everything. The second reason is that expanding the categories would increase the complexity of both the model and the research. By increasing categories, the model would have to train on more data and have a near impossible time separating between different group, and also with the large number of photos that exist in this dataset it would be nearly impossible to have groups for each photo. There would be a large amount of overlap as photos could fall into different groups. True this does occur with these three labeling groups, but the overlap is quite minimal.

Additional groups were added during different tests. These groups being images that were labeled as “Violence” and “Political”. The images labeled Violence contained very graphic images and the Political images contained scenes of public political figures. However, this idea was scrapped as it proved to complex and the overlap between groups became to great. These could be explored in future iterations of this research. The figures below demonstrate images from the ISIS Twitter dataset that fall into the assigned groups. To conclude, the three groups that are being used to group phots are Military, News, and Religion.

Figure 2 Example Military Photos

Figure 3 Example Religion Photos

Figure 4 Example News Photos

Initial Approach

As previously stated in this paper, the main objective of this research is to create a model that can predict the subject matter that is being depicted in the photos posted on Twitter. The initial approach involved using the links that were associated with each photo. Each tweet came with two different types of links that could be followed. The first would be a link that led directly to the tweet if it was available. This link would sometimes be followed to try and translate the tweet and provide some more context. Utilizing the text from the tweet and the image embedded in the tweet could be a possible future avenue of approach for a project. However, for the purpose of this first approach the only thing of concern for this project was the photo by itself devoid of the context of the tweet so the link that was used was the link directly to the photo. For all training and testing in this research, Python was the language of choice, and the Google Colab coding environment was used. The model that will be described in the next section relied on the use of links to photos.

Convolutional Neural Network

Both models that will be discussed in the following sections are both fundamentally Convolutional Neural Networks (CNN). A CNN utilizes the idea of deep learning to create neural networks that are commonly used in image classification. Deep learning is essence, is an attempt to make a machine learning model operate like a human brain. To explain, regarding images, humans can break down images and see different parts of an image and generate a sort of pattern. For example, if we see a bunch of sand and water and palm trees, we know we're looking at a beach, but a desert may just have the sand. Us as humans can see this unique difference in patterns and understand the little difference and deep learning embodies that idea. A CNN takes deep learning and breaks down images into a variety of different components by using matrices with different values to assign to different colors, or pixels, or already identified patterns (Saha, 2018). The way of creating these patterns is through the convolutional layers. These convolutional layers build on top of one another each one learning from the last.

Initial Model Approach

The first model that was used for attempting to train a model to predict the labels of images was the TensorFlow and Keras libraries. Specifically, within those packages the first model generated used the Sequential Model. A Sequential Model in Keras allows for the creation of a Convolutional Neural Network (CNN) one layer at a time. The layers can be chosen by the user and is built one layer at a time. Below in Figure 5, the breakdown of the different layers used by the model can be found and the table below it provides a description of each layer.

```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 sequential_1 (Sequential)   (None, 180, 180, 3)       0

 rescaling_2 (Rescaling)     (None, 180, 180, 3)       0

 conv2d_3 (Conv2D)           (None, 180, 180, 16)      448

 max_pooling2d_3 (MaxPooling  (None, 90, 90, 16)       0
 2D)

 conv2d_4 (Conv2D)           (None, 90, 90, 32)        4640

 max_pooling2d_4 (MaxPooling  (None, 45, 45, 32)       0
 2D)

 conv2d_5 (Conv2D)           (None, 45, 45, 64)        18496

 max_pooling2d_5 (MaxPooling  (None, 22, 22, 64)       0
 2D)

 dropout (Dropout)           (None, 22, 22, 64)        0

 flatten_1 (Flatten)         (None, 30976)             0

 dense_2 (Dense)             (None, 128)               3965056

 dense_3 (Dense)             (None, 3)                 387

=================================================================
Total params: 3,989,027
Trainable params: 3,989,027
Non-trainable params: 0
_________________________________________________________________
```

Figure 5 Model Summary

Table 3 Explanation of Layers

Layer	Layer Description
Sequential	Base layer identifying the type of model to be used.
Rescaling	Performed data augmentation that involved rotating and augmenting the coloring of the training data.
Conv2d	“This layer creates a convolution kernel that is convolved with the layer input to produce a tensor of outputs”(Module, 2022).
Max_Pooling	“Down samples the input along its spatial dimensions (height and width) by taking the maximum value over an input window for each channel of the input” (Module, 2022).
Dropout	Randomly drops out (by setting the activation to zero) several output units from the layer during the training process (Module, 2022).
Dense	Layer where all inputs are the outputs from the previous layer. Utilizes all inputs.

The Sequential Model was trained using the Ground Truth dataset. The model trained on 992 images gathered via the Selenium package and had the following split: 331 Military photos, 262 News photos, and 399 Religion photos. In addition to training on this dataset, Keras allows for two other means of improving the model, Data Augmentation and Dropout. Data Augmentation is a common approach in image classification and involves adding training data by taking each photo and making slight alterations to the image by either rotating the image or changing the coloring slightly. The other is dropout, which takes a percentage of the outputs from a layer and remove them. This creates a small series of neural networks as opposed to one large one. Though this model was quite easy to tune and to train on, testing is where problems occurred with this approach. To this end, there are no results to show for the ISIS dataset for this model at this time.

Problems with Initial Approach

There were two problems with the Tensor Flow image classification model that were not effectiveness based, but mainly user interaction based. By this, I simply mean that it was not conducive for inputting new testing data as the model that was chosen for the final approach. A user would have to go through and attach an ID to an image and then follow the link to get to the image. The best way to do a large number of photos was to loop through following the links, but it became a long process to even just process ten photos. This was tedious and was probably an avoidable problem, however, the work involved in doing this was not worth it as other models were easier to track and see how the model labeled images and input a large amount of the images. The second is that some URLs would lead back to the same image as someone may have simply retweeted it and the image appeared again. This made it hard as multiple images were being tested on multiple times, as opposed to a multitude of different unique testing images.

Ultimately, the Keras Sequential Model showed a lot of promise and may have proven to be a more ideal model. However, for the sake of this research this model was abandoned as it did not prove to be quite useful for the testing of new data. This model should not be summarily dismissed and may prove to be useful in future iterations of this research or for projects of a similar goal.

Final Approach

After the TensorFlow Keras Sequential Model proved to be inadequate for this project, a new approach was decided on. This final approach is broken into two separate sections as the final model that was used is the same, however, as will be discussed in the results section, there are two different approaches to training the model. In one iteration of the final model, the model was trained on the Ground Truth dataset, and another was trained on images from the ISIS Twitter dataset. However, the process of inputting these datasets into the model for training was an identical process.

Final Model Approach

The final model used to approach this research involved using the PyTorch library to implement another Convolutional Neural Network. However, this model provided a much more user-friendly approach to model building. Firstly, the data could be loaded in with relative ease. As opposed to going in and grabbing each link and following the link to the required image, all that was required was just a folder of images to unzip for both the training and testing set. Secondly it was quite easy to tune the model and switch between different types of models to use. The general model that was used is a CNN, however, under the umbrella term of CNN there are a lot of different implementations of a CNN. In the case of this research, both AlexNet, ResNet 18, and ResNet 34 were used to try and test this model. However, there are dozens of different models that can be used but these provided two of the most high-end models and allowed the opportunity to see how different types of models effected results.

Both ResNet-18 and ResNet-34 were developed by researchers at Microsoft. The goal of the ResNet models were to solve the degradation problems that exist with Deep Convolutional Neural Networks. To elaborate, the degradation problem is that with deeper networks, as the depths increase, the accuracy of the model eventually reaches a point where the accuracy starts to degrade quite quickly (He et al., 2016). The way that ResNet attempted to solve this problem is by letting these layers fit a residual mapping, as opposed to just stacking layers and hoping they fit (He et al., 2016)). Below in Figure 6, a simplified view of the architecture for all the various types of ResNets can be found. There are ResNets that have greater layers than 34, however, these were not used as they may have proven too complex and would slow down testing. In addition to testing on ResNet 18, AlexNet was used to try a completely new model.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 6 ResNet Architecture (He et al., 2016)

AlexNet is another attempt to improve the convolutional neural network that was created by an employee at Google. AlexNet attempts to try and improve performance by focusing on using both model and data parallelization. Model parallelization involves using different workers to train different parts of the model, while data dimensions use different workers to train on different examples of the data set (Krizhevsky, 2014). A more simplified version of this approach can be seen in Figure 7.

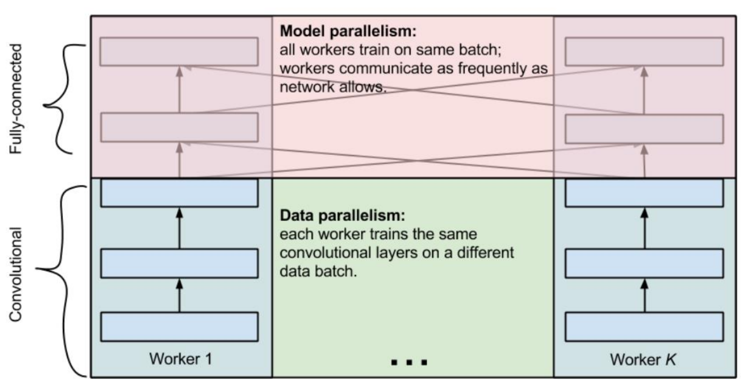

Figure 7 Simple AlexNet Architecture (Krizhevsky, 2014)

With PyTorch, a variety of other models exist that could have been tested on and the utilization of these will be discussed in this paper. However, the goal of this research was to at least find a model that could demonstrate good results. Utilizing other models and other libraries may prove to be more effective for this research and may prove to provide even more insight.

However, there are some constants across all the models that were used in training these models. First off, all PyTorch models utilized the Cross Entropy Loss function. What this function does is it looks at how far away the prediction probability is away from 0 or 1. The greater the distance, the higher the loss, as this is a logarithmic function. The goal when using this function is to reduce the loss and ultimately adjust the weights for the models while training. In addition to this, all models use stochastic gradient decent to help improve model performance over time by attempting to reduce the error of the model and reassigns weights to attempt to optimize (). The parameters for this function included an implementation of a learning rate of .001 and applying momentum at a value of .9. Adding momentum to SGD prevents the model from bouncing around specific weights and keep heading straight towards a desired optimum (). Then finally, the learning rate is decreased by .1 for every three epochs of the model.

When it came to tuning the model and finding the best method for model generation, a few parameters were tuned. The first, which has already been discussed, was trying to use different models. As previously stated, PyTorch comes with a wide variety of different models at its disposal that are easy to plug and play with. Secondly, the number of epochs used for each model was changed. Epochs is the number of times the training data is passed through the training model. For models testing on the Ground Truth dataset, epoch values of 5, 10, and 15 were used for testing as the dataset was much larger and could potentially timeout and epoch values of 10, 20, and 30 were used for the models trained on images from the ISIS Twitter data set as this dataset was much smaller for training and testing. Finally, the option to randomly rotate (rotated 15 degrees) and perform a horizontal flip on the data was either used or not used when testing. This was to see if adding a bit more variety to the data by causing some images to be randomly tilted or flipped horizontally would improve performance.

Chapter 4

Analysis of Results

In the following sections the results from model training and testing will be displayed, as well as an explanation of what data is being tested on. The metrics that are covered in the tables below are Accuracy, F1 score and precision. These three methods were chosen as they are a simple, quick way to see performance, as well as give us a good method of simply seeing how correct the model was overall in regard to different labels and general performance.

Results for Training on Ground Truth Data

The first model using the PyTorch method utilized Ground Truth data for training and tested on images from the ISIS Twitter dataset. In total, the training set consisted of 318 Military photos, 308 News photos, and 337 photos for Religion for a total of 963 photos that are both .png and .jpgs. As a reminder, this data came from data collected using the Selenium library allowing for the collection of photos from Google Images. There was slightly more phots in Religion than any other group because there was a bit more variety in that group. In regard to testing, there was 190 randomly selected photos that had already been labeled and had a break down of 60 Military photos, 55 News photos, and 75 Religion photos. The training and testing were split to have an 80-20 split. This was a little skewed towards Religion as this label appeared a bit more often than other labels. In the following tables, some statistics are broken down by labels and those labels have been shortened to M (Military), N (News), and R (Religion).

Table 4 Ground Truth Performance with Rotation and Horizontal Flip

ResNet 18	Accuracy	Accuracy (M)	Accuracy (N)	Accuracy (R)	F1 Score	F1 Score (M)	F1 Score (N)	F1 Score (R)	Precision (M)	Precision (N)	Precision (R)
5 Epochs	83.68%	93.33%	72.73%	84.00%	83.51%	89.60%	76.19%	84.00%	86.15%	80%	84%
10 Epochs	83.68%	93.33%	72.73%	84.00%	83.51%	89.60%	76.19%	84.00%	86.15%	80%	84%
15 Epochs	83.16%	90.00%	83.64%	77%	83.20%	90.00%	78.63%	81.11%	90%	74.19%	85.29%
ResNet 34	**Accuracy**	**Accuracy (M)**	**Accuracy (N)**	**Accuracy (R)**	**F1 Score**	**F1 Score (M)**	**F1 Score (N)**	**F1 Score (R)**	**Precision (M)**	**Precision (N)**	**Precision (R)**
5 Epochs	82.63%	95.00%	90.91%	66.67%	82.50%	91.20%	78.13%	78.74%	87.69%	68.49%	96.15%
10 Epochs	78.42%	80.00%	85.45%	72.00%	78.64%	85.71%	75.20%	75.52%	92.31%	67.14%	79.41%
15 Epochs	81.58%	95.00%	87.27%	67%	81.34%	88.37%	78.05%	78.13%	82.60%	70.59%	94.34%
AlexNet	**Accuracy**	**Accuracy (M)**	**Accuracy (N)**	**Accuracy (R)**	**F1 Score**	**F1 Score (M)**	**F1 Score (N)**	**F1 Score (R)**	**Precision (M)**	**Precision (N)**	**Precision (R)**
5 Epochs	72.11%	78.33%	78.18%	62.67%	71.99%	78.33%	71.67%	67.14%	78.33%	66.15%	72.31%
10 Epochs	72.11%	78.33%	78.18%	62.67%	71.99%	78.33%	71.67%	67.14%	78.33%	66.15%	72.31%
15 Epochs	72.11%	78.33%	78.18%	62.67%	71.99%	78.33%	71.67%	67.14%	78.33%	66.15%	72.31%

Table 5 Ground Truth Performance without Rotation and Horizontal Flip

ResNet 18	Accuracy	Accuracy (M)	Accuracy (N)	Accuracy (R)	F1 Score	F1 Score (M)	F1 Score (N)	F1 Score (R)	Precision (M)	Precision (N)	Precision (R)
5 Epochs	84.21%	91.67%	69.09%	89.33%	83.96%	89.43%	77.55%	84.28%	87.30%	79.76%	88.37%
10 Epochs	82.63%	90.00%	74.55%	82.67%	82.57%	89.43%	77.55%	84.27%	87.10%	82.67%	77.36%
15 Epochs	84.21%	91.67%	69.09%	89.33%	83.96%	89.43%	77.55%	84.28%	87.30%	79.76%	88.37%

ResNet 34	Accuracy	Accuracy (M)	Accuracy (N)	Accuracy (R)	F1 Score	F1 Score (M)	F1 Score (N)	F1 Score (R)	Precision (M)	Precision (N)	Precision (R)
5 Epochs	83.16%	90.00%	78.18%	81.33%	83.10%	87.10%	78.90%	82.99%	84.38%	79.63%	84.72%
10 Epochs	83.16%	90.00%	83.63%	77.33%	83.16%	88.52%	79.31%	81.69%	87.10%	75.41%	86.57%
15 Epochs	81.58%	90.00%	89.09%	89.33%	81.69%	90.76%	86.56%	78.20%	91.53%	67.12%	89.66%
AlexNet	**Accuracy**	**Accuracy (M)**	**Accuracy (N)**	**Accuracy (R)**	**F1 Score**	**F1 Score (M)**	**F1 Score (N)**	**F1 Score (R)**	**Precision (M)**	**Precision (N)**	**Precision (R)**
5 Epochs	80.53%	81.67%	78.18%	81%	80.56%	80.99%	81.90%	79.22%	80.33%	86.00%	79.22%
10 Epochs	80.53%	81.67%	78.18%	81%	80.56%	80.99%	81.90%	79.22%	80.33%	86.00%	79.22%
15 Epochs	80.53%	81.67%	78.18%	81%	80.56%	80.99%	81.90%	79.22%	80.33%	86.00%	79.22%

Based on these results there are a few observations that can be made. Firstly, ResNet 18 performed better generally than the other models. In both sets of testing the performance in all general categories (not for specific labels) performed a bit better with a general accuracy of 83% for both sets of testing, while ResNet 34 usually dipped by about 1% to 2% and AlexNet always performed worse than the other two. AlexNet, it appears, had no improvement when more epochs were added, and simply stayed stagnant at its predicate capabilities. In general, it also appears that the worse performing label was the News label. This had scores that typically performed worse than the other labels. Finally, it appears that there isn’t a clear correlation from this data that shows that the horizontal flip and rotation have any bearing on performance.

Results for Training on ISIS Dataset

The second iteration of the PyTorch model both trained and tested on photos contained within the ISIS Twitter dataset. The training dataset only contained a total of 285 photos with the following breakdown: 114 Military photos, 87 News, and 84 Religion. The testing set contained 72 photos with a split of 24 across each label. A new round of labeling was done on these photos by all researchers as opposed to one. There are two reasons that the size of the dataset is much smaller in this iteration of the model. Firstly, these photos are much larger in size than the photos scraped from Google. Google photos were only a few kilobytes, while the images from the ISIS Twitter dataset were much closer to one full megabyte. Secondly, since these photos were from the same dataset it was assumed that a lot of data would be needed to create a good model and the results show this. As a final note, less data was also used as labeling photos manually was a very time-consuming procedure.

Table 6 ISIS Twitter Dataset Performance with Rotation and Horizontal Flip

ResNet 18	**Accuracy**	**Accuracy (M)**	**Accuracy (N)**	**Accuracy (R)**	**F1 Score**	**F1 Score (M)**	**F1 Score (N)**	**F1 Score (R)**	**Precision (M)**	**Precision (N)**	**Precision (R)**
10 Epochs	93.06%	100.00%	91.67%	87.50%	93.01%	94.12%	93.61%	91.30%	88.89%	95.65%	95.45%
20 Epochs	93.06%	95.83%	87.50%	95.83%	93.02%	95.83%	89%	93.88%	95.83%	91.30%	92%
30 Epochs	91.67%	95.83%	87.50%	92%	91.64%	93.88%	89.36%	91.67%	92.00%	91.30%	91.67%
ResNet 34	**Accuracy**	**Accuracy (M)**	**Accuracy (N)**	**Accuracy (R)**	**F1 Score**	**F1 Score (M)**	**F1 Score (N)**	**F1 Score (R)**	**Precision (M)**	**Precision (N)**	**Precision (R)**
10 Epochs	94.44%	87.50%	95.83%	100.00%	94.40%	93.33%	93.88%	96.00%	100.00%	92.00%	92.31%
20 Epochs	94.44%	91.67%	91.67%	100.00%	94.41%	93.61%	96.00%	95.65%	95.65%	92.30%	92.31%

30 Epochs	94.44%	91.67%	91.67%	100%	94.41%	93.62%	93.62%	96.00%	95.65%	95.65%	92.31%
AlexNet	**Accuracy**	**Accuracy (M)**	**Accuracy (N)**	**Accuracy (R)**	**F1 Score**	**F1 Score (M)**	**F1 Score (N)**	**F1 Score (R)**	**Precision (M)**	**Precision (N)**	**Precision (R)**
10 Epochs	83.33%	87.50%	87.50%	75%	83.36%	80.77%	87.50%	81.82%	75.00%	87.50%	90.00%
20 Epochs	83.33%	87.50%	87.50%	75%	83.36%	80.77%	87.50%	81.82%	75.00%	87.50%	90.00%
30 Epochs	83.33%	87.50%	87.50%	75%	83.36%	80.77%	87.50%	81.82%	75.00%	87.50%	90.00%

Table 7 ISIS Twitter Dataset Performance without Rotation and Horizontal Flip

ResNet 18	**Accuracy**	**Accuracy (M)**	**Accuracy (N)**	**Accuracy (R)**	**F1 Score**	**F1 Score (M)**	**F1 Score (N)**	**F1 Score (R)**	**Precision (M)**	**Precision (N)**	**Precision (R)**
10 Epochs	95.83%	100.00%	91.67%	95.83%	95.83%	100.00%	93.62%	93.88%	100.00%	95.65%	92.00%
20 Epochs	93.06%	95.83%	91.67%	91.67%	93.06%	95.83%	92%	91.67%	93.06%	95.83%	92%
30 Epochs	91.67%	91.67%	91.67%	92%	91.77%	95.67%	88.00%	91.67%	100.00%	84.62%	91.67%
ResNet 34	**Accuracy**	**Accuracy (M)**	**Accuracy (N)**	**Accuracy (R)**	**F1 Score**	**F1 Score (M)**	**F1 Score (N)**	**F1 Score (R)**	**Precision (M)**	**Precision (N)**	**Precision (R)**
10 Epochs	93.06%	91.67%	93.88%	91.67%	93.05%	93.62%	93.88%	91.67%	95.65%	92.00%	91.67%
20 Epochs	90.28%	79.17%	95.83%	95.83%	90.12%	86.36%	92.00%	92.00%	95.00%	88.46%	88.46%
30 Epochs	90.28%	83.33%	91.67%	95.83%	90.32%	88.89%	93.62%	88.46%	95.24%	95.65%	82.14%
AlexNet	**Accuracy**	**Accuracy (M)**	**Accuracy (N)**	**Accuracy (R)**	**F1 Score**	**F1 Score (M)**	**F1 Score (N)**	**F1 Score (R)**	**Precision (M)**	**Precision (N)**	**Precision (R)**
10 Epochs	90.28%	87.50%	87.50%	96%	90.39%	89.36%	93.33%	88.46%	91.30%	100.00%	82.14%
20 Epochs	90.28%	87.50%	87.50%	96%	90.39%	89.36%	93.33%	88.46%	91.30%	100.00%	82.14%
30 Epochs	90.28%	87.50%	87.50%	96%	90.39%	89.36%	93.33%	88.46%	91.30%	100.00%	82.14%

The clearest observation to be made from these results is that using the ISIS Twitter dataset provides high accuracy. This is unsurprising as the model was trained and tested on subsets from the same larger dataset. These results are valuable, though this model should not be generalized for all media of the same type. This model would most likely perform poorly when tested on data from another terrorist group. Secondly, AlexNet still replicates results regardless of the number of epochs used. At this time, no explanation can be provided as to why this occurs. Finally, the best performing model over all in this model would be the ResNet 34 with Rotation and Horizontal Flip, as the overall accuracy is around 94% for all iterations, though the best single iteration of the model occurred at 10 epochs with ResNet 18 and no Rotation and Horizontal flip with an accuracy of 95.83%.

Final Summary of Results

Based on the results for both sets of testing it seems as if the models generally performed well when it came to labeling photos that were considered “Military”, which usually had accuracy around the 90% range. This makes sense as these photos are more easily identified. They contain a lot of unique imagery such as camouflage, military vehicles, weapons, and other aspects that would make them more easily identifiable. This is in juxtaposition to the other labels of Religion and News which had a tendency to perform much worse. News was a little more surprising, as a lot of the photos that were tested on contained a red bar across the bottom with scrolling text that is common in news broadcasts. News performed especially poorly in the Ground Truth trained models and perhaps and increase in images with that distinct red bar would increase performance. With Religion, the poor performance can probably be attributed to a lack of variety in clear images. For example, a lot of the religious photos had an image of a mosque in them, but mosques are building that are prominent in this area and could easily appear in both News and Military photos. An increase in data in different unique symbols for religion would probably improve model performance.

Chapter 5

Limitations and Areas of Improvements

The following sections outline what caused limitations in the ability to improve the research and provide more significant results for use at a larger scale and areas of improvement where the research itself could have been conducted in a more improved manor.

Limitations

One of the major limiting factors in this research is the actual hardware being used to run the programs. Although the Python program used to generate models was run in a cloud environment, it still had to utilize the resources available to the computer it was on. The main limiting factor in my computer was the processor that severely limited how quickly tasks could be completed and how much data could be analyzed at a time. On average models took an hour and half to complete testing. Having more processors would allow for the use of even more workers. There were very few options if any for more improved computers, but with how much data I kept creating and recreating on this computer it made sense to just continue working with the computer I have now. Using a more powerful computer would allow for the use of more complex models and can allow for even more data to create an even stronger model.

The second big limitation on this project was the number of people working on this research. One of the key aspects of the data that couldn’t be explored is the actual text that some images had and some of the text attached to the tweets. In total there were three people working on this project and only one of us had a slight form of understanding of Arabic. Obviously there exist options out there to translate words, however, these might be missing the true meaning of things. There may be more subtle nuances between the use of words in different causes that Natural Language Processing may not be able to capture. More informative research in this regard could utilize a couple of researchers who are fluent in Arabic and could sit and label a lot of different photos and provide more insight into the image. Also, more researchers who are labeling photos could help avoid bias in labeling photos and even the categories that were used for labeling. People from different departments or people who are of Middle Eastern decent might have a better understanding of the subject of the photos and what labeling should actually be used. Maybe something that are labeled as religious may have symbology that is not actually religious but more of a cultural subject. However, getting more researchers is not easily done and may have created more complications.

Areas of Improvement

First and foremost an area that could have improved the training of the model would be tuning a different set of parameters. Each model used had its own means of optimizing the model and tuning on these would have been more beneficial. The default parameters for each model were used for testing. In order to potentially get even more accurate models then the results depicted. Also, there were a lot more parameters that could have been tuned that were outside of the type of model used. For example, different learning rates could have been tested and momentum could have been removed. A greater focus should have been put on analyzing these values as they would have been applied to all models and could have been more crucial to altering results and improving the models. In addition to this, a wider variety of models could have been used. PyTorch offers more models than different layers of ResNet and Alex net, it also offers VGG, SqueezeNet, DenseNet, and perhaps a dozen more state of the art models that exist and have been researched extensively. One of these models may have had an even better performance or been more optimal for this project. More research should have been put into these models. Finally, regarding the labels used for the dataset, more general terms should have been used to identify images. The label that may have been too specific was the Military label. A traditional thought of Military images may be of tanks and guns and soldiers, which were included in this group for training. However, some other things like destruction and people bleeding were used in this labeling group as well. These images were most likely the result of military action; however, this may not have been the case. A more general term like violence or war may have been more accurate. Improving the labeling system would prevent someone who's looking at a labeled image for the first time from getting stuck in one way of thinking about an image.

Chapter 6

Future Work and Conclusion

The discoveries outlined in this thesis are a simple method in which to analyze ISIS media content. However, as will be outlined below, there is much more research that can be done in this field.

Future Work

Let his research serve as just the tip of the iceberg for what interesting finds can be made utilizing this dataset or this concept of social media research. Just for this ISIS Twitter dataset there are a multitude of avenues for new approach. One of these new approaches would be looking into not just the photo content, but also the gif and videos that exist. This type of media was avoided to provided simplicity, but these were left untouched and may uncover more insight into what type of media ISIS uses and what the subject of this content is. In addition to just the different types of media, there is also the tweets themselves that are attached to each photo. These can be used as a method of providing more context to the photo and assist in labeling the image. The labels themselves can be expanded on to include a more comprehensive view of all the possible images exposed by ISIS and ISIS sympathizers. More groups may provide more narrowed down observations for each photo, or less labels can improve the generalizability of the model. Finally, these models can be tested on different terrorist groups from all parts of the world to see how the models stack up when used on different groups.

Conclusion

From what has been discussed in the above sections, there is still a lot of research that can be done and should be done as it comes to this dataset. However, goal of this research is to provide proof that a method for identifying the photos from an extremist group like ISIS is possible. The benefits for this can be huge as they could provide a method for social media sites to have an improved method of identifying hazardous content and a method for our own defense groups to grasp a better understanding of groups that pose a threat to us.

BIBLIOGRAPHY

Benigni, M. C., Joseph, K., & Carley, K. M. (2017). Online extremism and the communities that sustain it: Detecting the ISIS supporting community on twitter. PLoS One, 12(12) doi:http://dx.doi.org.ezaccess.libraries.psu.edu/10.1371/journal.pone.0181405

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Homepage. (n.d.). Retrieved April 18, 2021, from http://www.casos.cs.cmu.edu/

Krizhevsky, A. (2014). One weird trick for parallelizing convolutional neural networks. arXiv preprint arXiv:1404.5997.

Magdy, W., Darwish, K., & Weber, I. (2016). FailedRevolutions: Using twitter to study the antecedents of ISIS support. First Monday, https://doi.org/10.5210/fm.v21i2.6372

Mansour, S. (2018). Social media analysis of user's responses to terrorism using sentiment analysis and text mining. Procedia Computer Science, 140, 95-103.

Module: Tf.keras.layers | TensorFlow Core v2.8.0. (n.d.). Retrieved March 27, 2022, from https://www.TensorFlow.org/api_docs/python/tf/keras/layers

Ottolenghi, E. (2021, March 26). Social media is an Intel gold MINE. why aren't governments using it? Retrieved April 12, 2021, from https://foreignpolicy.com/2021/03/26/social-media-big-tech-facebook-twitter-intelligence-sharing-law-enforcement/

Pandith, F., Adjunct senior fellow at the Council on Foreign Relations, Ware, J., & Research associate for counterterrorism at the Council on Foreign Relations. (2021, March 22). Teen terrorism inspired by social media is on the rise. here's what we need to do. Retrieved April 12, 2021, from https://www.nbcnews.com/think/opinion/teen-terrorism-inspired-social-media-rise-here-s-what-we-ncna1261307

Saha, S. (2018, December 17). A Comprehensive Guide to Convolutional Neural Networks— The ELI5 way. Medium. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Siegel, A. A., & Tucker, J. A. (2018). The Islamic State's information warfare: Measuring the success of ISIS's online strategy. Journal of Language & Politics, 17(2), 258–280. https://doi-org.ezaccess.libraries.psu.edu/10.1075/jlp.17005.sie

Wright, S., Denney, D., Pinkerton, A., Jansen, V. A. A., & Bryden, J. (2016). Resurgent Insurgents: Quantitative Research Into Jihadists Who Get Suspended but Return on Twitter. Journal of Terrorism Research, 7(2), 1–13. DOI: http://doi.org/10.15664/jtr.1213

ACADEMIC VITA

Michael J. Fanelli

Education:

The Pennsylvania State University (PSU) – University Park, PA   Expected Graduation: May 2022

(BS) – Applied Data Science (w/Minor: Security and Risk Analysis) - College of Information Science and Technology (IST) Schreyer Honors College

Skills/Relevant Coursework Summary

- Calculus with Analytic Geometry / Statistical Data Analysis / Basic Discrete Mathematics
- Experienced with SQL / NoSQL / MongoDB
- Proficient with Python, Experience with Linux / Windows Command Line / Spark /R / GitHub / KML
- Created and deployed different Machine Learning algorithms and Data Visualizations (scikit-learn, d3, TensorFlow, PyTorch, plotly)
- Excellent Communication Skills / Organization Skills/Project Management Skills, Experienced with Microsoft Suite
- Knowledge of Network Security concepts, trends, and techniques, Experience and understanding of Project/Software development Lifecycle (Agile, Waterfall, etc.)
- Experience with Parallel Computing

Awards and Honors

- Meyer Family Honors Scholarship – 2020, 2021
- IST Dean's List – Fall 2018, Spring 2019, Fall 2019, Spring 2020, Fall 2020, Spring 2021, and Fall 2022
- Security+ Certified – July 2021
- Active Secret Government Clearance

Employment/Research History:

Penn State University Applied Research Lab (CINO Group) - May 2021 to Present:

- Worked in an offsite/onsite location performing a variety of tasks including product testing/deployment, data analysis/visualization, and report generation (bug reports, product set up reports, etc.). All of these in alignment with assisting in the goals of the group we were contracted to work for.

Penn State University Applied Research Lab (PMO Group) – January 2020 to May 2021

- As an intern for the Project Management Office (PMO) at the Applied Research Lab (ARL) at Penn State, I worked alongside Project Managers (in-person and remotely) to apply business analysis and project management best practices in order to deliver real results to ARL. I worked in fields focusing on Technology Management, Security & Compliance, Research Enablement, and Business and Administration, all the while learning and expanding upon PMO processes and tools.

Leadership Experiences/Extra Curricular Activities:

THON Operations Committee Icebreaker Chair – October 2018 to February 2021

- Facilitated team bonding and communication activities that helped the group work more cooperatively and successfully by learning more about one another and how to trust one another.

IST Learning Assistant – August 2019 to December 2019

- Assist students taking IST 110 by conducting review sessions during weekly office hours; address course related questions and concerns; assist with test grading and as needed provide individual feedback; and collect/summarize feedback data for purposes of improving student learning and overall course presentation.

IST SLO Treasurer – May 2019 to May 2020

- Attend weekly SLO Executive Committee meetings. Collect dues from new and renewing members; organize and plan community and group events; track and provide accounting for all organizational funds; and facilitate payment of all bills and expenses.