

## **Project Topic: Detecting Malicious URLs**

**Students: Ashraful Tauhid, Surendra Pathak, Joshua Menefee**

**Team: 11**

- 1. Base Set of Papers/Articles:** In our last report, our group mentioned that we would be selecting a base set of articles and/or papers to review and analyze as guidance for our project's development. The papers/articles that we've decided on include:
  - a. Kumi, Sandra, et al. "Malicious URL Detection Based on Associative Classification." *Entropy*, vol. 23, no. 2, 2021, p. 182., <https://doi.org/10.3390/e23020182>.
  - b. Sahoo, Doyen, et al. Malicious URL Detection Using Machine Learning: A Survey, 21 Aug. 2019.
  - c. "Using Machine Learning to Detect Malicious Urls." KDnuggets, <https://www.kdnuggets.com/2016/10/machine-learning-detect-malicious-urls.html>.
  - d. "Malicious URL Detection." Malicious URL Detection, <http://athena.ecs.csus.edu/~narendrp/>.
- 2. Data Collection Progress:** In the past 2 weeks we have begun collecting URLs for our dataset, both through our group's own contributions, and through URLs that we've found through other datasets found online (Refer to our previous progress report on the different sources that we will potentially gather data from). As we collect our data, we are compiling it into a CSV file that will be read and loaded into our final model as the dataset.
- 3. Plans For Model Implementation:** As for the model itself, using the guidance of the documents we chose, we will most likely implement a Convolutional Neural Network that will classify URLs as either safe or malicious. As for the libraries we will use, a couple of our options include the WEKA library and SKLearn. As of right now we don't have anything finalized but we plan on deciding on a final design for our CNN's structure once we have more data collected and pre-processed.
- 4. Remaining Tasks:**
  - a. Continue using our selected documents for guidance on our project's development.
  - b. Continue collecting/pre-processing data for our final dataset.
  - c. Finish designing our CNN
    - i. Begin implementing our model after designing.
  - d. Analyze the results and compare them to the results found in our selected documents.
  - e. Begin compiling our findings into a final report/presentation.