Week 2 Direct Independent Study Progress

* **Topic Covering**: Al-Powered Phishing Detection Using NLP & Explainable Al.
* **Covered Paper: Phishing Detection Using NLP and Machine Learning**
  + **Link**: ["Phishing Detection Using NLP and Machine Learning" by Apurv Mittal, Dr Daniel Engels et al.](https://scholar.smu.edu/datasciencereview/vol6/iss2/14/)
  + **Citation:**  *Mittal, Apurv, Daniel Engels, Harsha Kommanapalli, Ravi Sivaraman, and Taifur Chowdhury. "Phishing Detection Using NLP and Machine Learning." SMU Data Science Review 6, no. 2 (2022): Article 14. https://scholar.smu.edu/datasciencereview/vol6/iss2/14.*
  + **Notes**: The authors of this paper introduce the DARTH framework, which uses Natural Language Processing (NLP) and machine learning techniques to detect phishing emails. The structure analyzes various features of emails, such as body text and metadata, using multiple models. By combining the results of these analyses, the system classifies emails as malicious or legitimate. The study utilizes a dataset of over 150,000 emails, achieving a precision of 99.97% and an F-score of 99.98%.
  + **Covered Paper**: **A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME**
  + **Link**: [[2305.02012] A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME](https://arxiv.org/abs/2305.02012)
  + **Citation**: Salih, Ahmed, Zahra Raisi-Estabragh, Ilaria Boscolo Galazzo, Petia Radeva, Steffen E. Petersen, Gloria Menegaz, and Karim Lekadir. "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME." arXiv preprint arXiv:2305.02012, May 2023. https://arxiv.org/abs/2305.02012.
  + **Notes**: The authors of this paper give a comprehensive analysis of two prominent Explainable AI (XAI) methods: SHAP and LIME. They discuss how these methods create explainability metrics and propose a framework for interpreting their results. The study also showcases the strengths and weaknesses of SHAP and LIME, particularly in the presence of feature collinearity and model dependency.

**Relevance of Paper to my Study**

* The first paper highlights the effectiveness of combining multiple models for enhanced accuracy.
* It offers insights into feature extraction and analysis relevant to phishing emails.
* The second paper gives a deep understanding of SHAP and LIME, aiding in the selection of appropriate XAI methods.
* It gives a foundation for interpreting model predictions in phishing detection systems.

# **Technical Setup**

* Using the Pycharm software with Python 3.11 being used as the environment.
* Installed needed libraries such as:
  + sckit-learn
  + Matplotlib
  + Seaborn
  + Jupyterlab
  + Torch
* Dataset to be used can be retrieved from:
  + PhishTank ([PhishTank > Developer Information](https://www.phishtank.com/developer_info.php))
  + Ember ([GitHub - elastic/ember: Elastic Malware Benchmark for Empowering Researchers](https://github.com/elastic/ember))
  + Kaggle ([Search | Kaggle](https://www.kaggle.com/search?q=phishing))
  + SpamAssassin (Note: This is great for legitimate email data) ([Index of /old/publiccorpus](https://spamassassin.apache.org/old/publiccorpus/))