**Project Motivations**

**Team Members:** Dauren, Himangshu, Satvik, Tigran

**What is the purpose of your project? What will you achieve?**

This project's goal is to develop a novel prediction churn reduction method. Prediction churn is the divergence in predictions made by a machine learning model after it has been retrained on new data. Churn needs to be minimized in order to keep model predictions more consistent for smooth operation of downstream tasks. Reduced churn also allows better insight into model performance and model behavior. Our results will also be made available to the client and the broader public through an open-source package.

We will gain a strong understanding of the weaknesses and areas for improvement in the existing prediction churn literature. It will also enable us in determining how to design a new algorithm based on past deficiencies identified in our literature review. We will also learn how to write industry-standard code in the development of the churn reduction package.

**Why is this project a valuable pursuit and who will it serve?**

The project is valuable since the techniques we propose will result in a significant improvement in topic prediction churn in general. The creation and characterization of model training algorithms that reduce "noisy" prediction churn is a nascent ML research field with potentially critical impact both to Proofpoint and to any industry which employs ML models at large.

As a result, it is intended that our work serve as a starting point for our client proof point, a leading global cybersecurity and compliance company, to implement/analyze the methodologies we propose, as well as for other researchers working on or planning to work on prediction churn.

**What similar work has already been done?**

Similar studies have been conducted in prediction churn using distillation techniques, locally adaptive label smoothing and the anchor method. The literature is limited on the topic of churn, however there are other bodies of work that deal with similar topics such as model instability and model confidence.