Chapter 9
Ensuring Privacy and Security in Machine Learning:
A Novel Approach to Efficient Data Removal

Velammal
Anna University, Chennai, India

N. Aarthy
Anna University, Chennai, India

ABSTRACT

Modern systems generate vast amounts of data, creating complex data networks. Users prioritize the safety, security, and privacy of their data. This project focuses on efficiently removing or erasing data from the machine learning model upon user request, addressing privacy concerns. Under GDPR, users can request the deletion of sensitive data from both user records and the machine learning model that has processed the data. Additionally, the project employs the SISA approach to address errors and attacks by dividing the dataset into shards and implementing a slice-based ensemble learning technique. Each shard functions as an independent model, and after training, a majority voting approach aggregates these models into a final model. Experimental results demonstrate reduced retraining costs, as only the remaining slices are retrained instead of the entire model.

INTRODUCTION

Machine Unlearning is a domain which is contradictory to Machine Learning. In Machine Learning, a machine will be made to learn the data set that is provided and produce a desired model with high accuracy whereas, in Machine Unlearning, machine will be made to unlearn the data that the model had previously learned. It is necessary for unlearning research (Ma,2022) to develop an algorithm that can take a trained machine-learning model as input and produce a new one without the requisite data. Retraining the model from scratch without the need to relearn the

DOI: 10.4018/979-8-3693-2964-1.ch009

training data is a fundamental tactic. Unlearning research aims to reduce the high computational cost associated with this. With the rising importance of data privacy, data is becoming a major concern. Data privacy is all about restricting the access to personal data which involves deciding who should not have access to it. Deletion of accounts simply delinks the data from database but for complete privacy the proposed model must ensure that the system forgets the data once and for all. Users desire systems to forget particular data for a variety of reasons, including its lineage. Users who are concerned about new privacy dangers in a system frequently want the system to forget their data and history. A detector must forget the injected data if an attacker taints it by adding manually generated data to the training set. This is necessary for the detector to regain security. A user can reduce noise and inaccurate items to ensure that a recommendation engine provides effective recommendations. Let's break down the problem in more detail so it is easy to see how it varies from other privacy definitions. The user-provided information that is asked to be unlearned is represented by the letter d. It is imperative to devise an unlearning strategy that, without starting from scratch, produces the same model distribution as retraining. In the naive approach, where the entire model is retrained after deletion, the computational cost is drastically high, and it takes a lot of time. Therefore, in order to better match with the objectives of new laws on privacy, an alternate strategy that investigates the deterministic definition is required. (Singh,2023) has devised docker swarm concept to load and to efficiently handle the data for bigdata applications. Here the model unlearns the user requested data. If the distributions do not match, there must be some influence on the system from them to account for the discrepancy. One way to think of an unlearning environment is as a probabilistic setting where most of the contribution of the user-requested data is eliminated, and an unlearning method only approximates the retraining distribution. On certain websites, users who knowingly or unknowingly provide personal information about themselves consent to the corporations running those platforms using that information for a range of purposes, such as selling it to marketers or using it to improve their prediction models. It becomes challenging for businesses to undo the impact of data acquired if a user decides not to let such information about them to be used by them, particularly if the data was used to train machine learning models. Most of the people, considering their privacy, don't want their data to be shared. For instance, consider a Man 'X' who want to delete his Account 'A'. He has just deleted his account by deactivating it. Here it should be understood that deleting an account doesn't mean that X's personal information is completely deleted. The Machine Model that is created using a dataset consisting of X's information still has a trace of X's details. But, it is not acceptable to have privacy sensitive data as a part of Machine learning model. Subsequently, in the situation of encountering errors and attacks, unlearning the incorrect or poisoned data sample is an obvious need. In elaborate, if a data administer, who collects and organizes data, has entered an incorrect sample and has let the machine learn the data, it results in the model which couldn't produce a desired output. Lastly, in the case of being trapped in the data attack called data pollution, there is a demand to erase the polluted data sample. In data pollution, the hacker tries to inject a malicious data sample to the training set causing the machine to learn the training set comprising the injected data sample. Removing or deleting the inserted data sample from the training set and relearning the clean set could be the solution.