US Laws on Privacy

- Title 13: Every person with access to Census data is sworn for life to protect your information and understands that the penalties for violating this law are applicable for a lifetime (federal prison sentence of up to five years, a fine of up to $250,000, or both).
- Title 26: Conditions under which the IRS may disclose Federal Tax Returns and Return Information (FTI) to other agencies, including Census Bureau.
- Health Insurance Portability and Accountability Act (HIPAA)
- Children's Online Privacy Protection Act (COPPA)
- Family Educational Rights and Privacy Act (FERPA)

Gangwal, A., Singh, S., & Srivastava, A. (2023, April). AutoSpill: Credential Leakage from Mobile Password Managers. In *Proceedings of the Thirteenth ACM Conference on Data and Application Security and Privacy* (pp. 39-47).

"The majority of popular Android PMs considered in our experiments were found vulnerable to AutoSpill;"

"IIIT-H alerted Google which acknowledged the security breach."

Thursday, October 19, 2023

DECCAN Chronicle

Home | Latest News | South | Cities | Nation | Sports | World | Entertainment | Opi

Home » Technology » Mobiles and Tabs » October 17, 2023

TECHNOLOGY

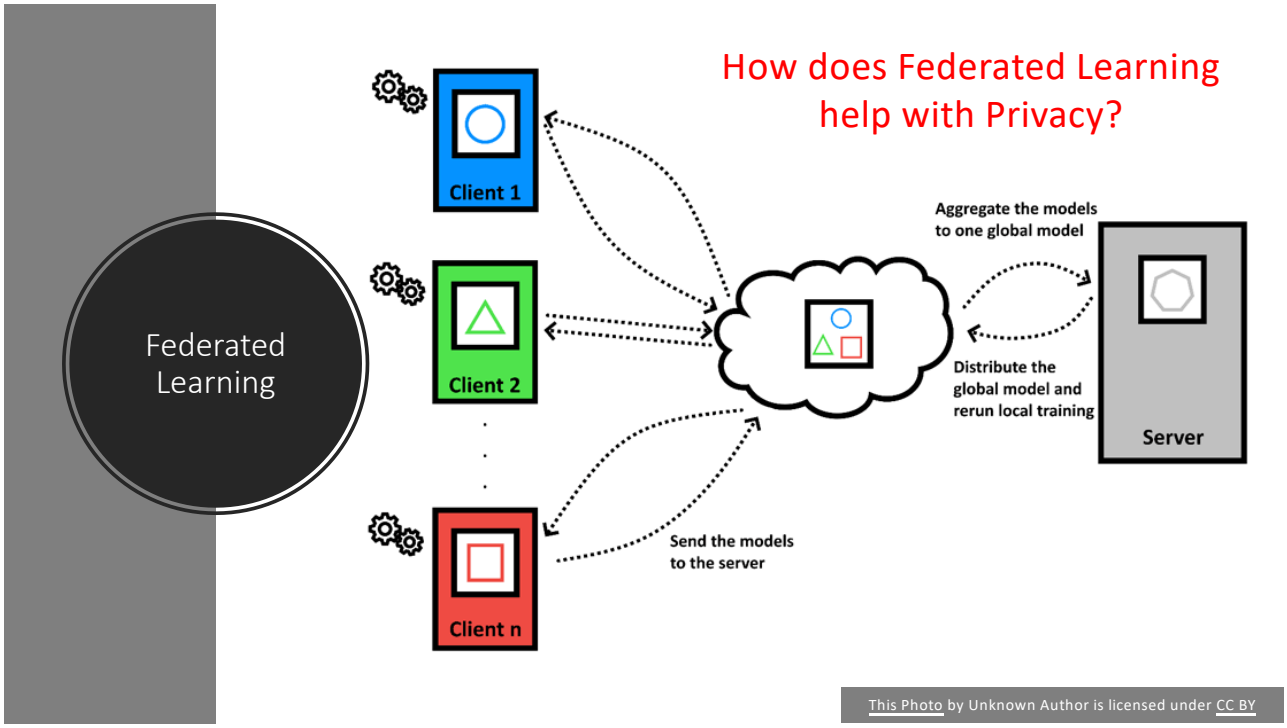
IIIT-H Spots Data Leak In Apps' Use of Autofill

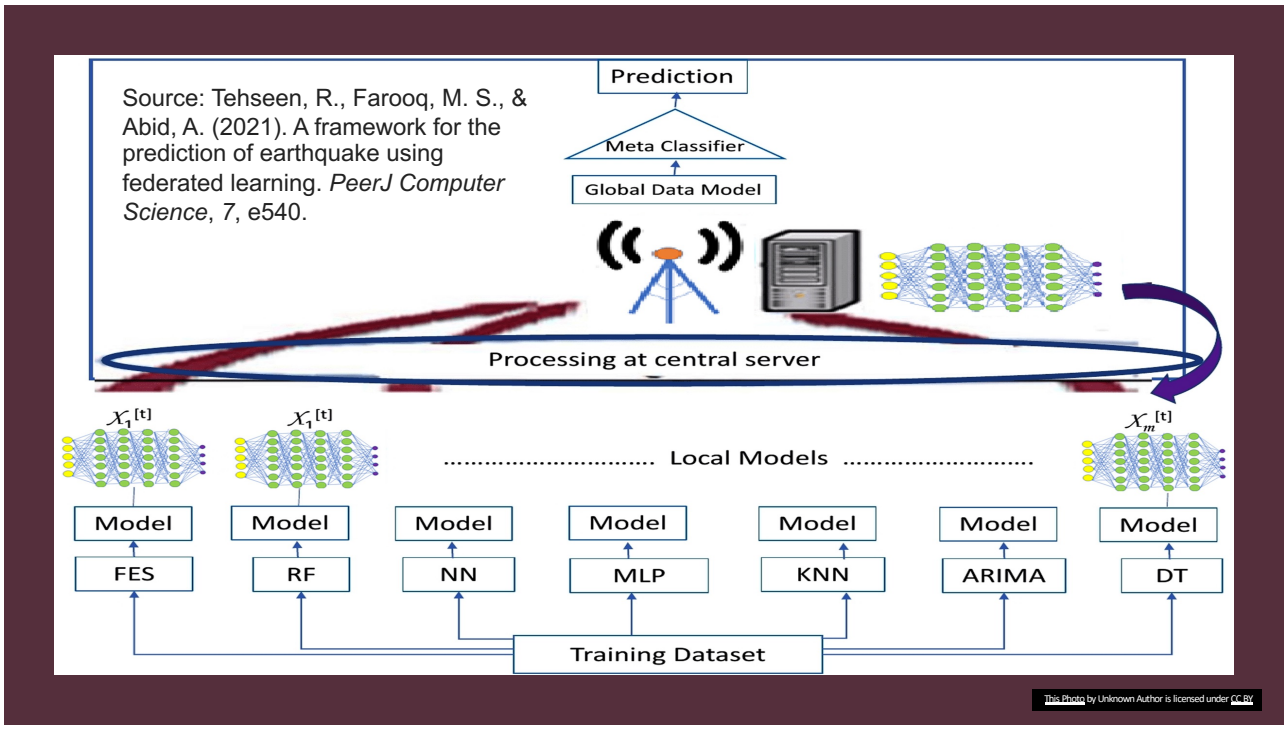
Deccan Chronicle.| DC Correspondent Published on: October 17, 2023 | Updated on: October 17, 2023

When a user tries to log into an app on the Android operating system (OS), the OS generates an autofill request to the password manager

The expected autofill behavior of PMs

How does Federated Learning help with Privacy?





Alleviating privacy concerns with SVM in federated learning

1. Differential privacy adds noise to the training process, which makes it more challenging to infer information about specific data points, including support vectors.
2. Federated learning systems often use secure aggregation techniques like federated averaging to aggregate model updates from multiple clients while preserving privacy.
3. Clients can use homomorphic encryption to send encrypted updates to the central server. The central server can perform computations on the encrypted data without decrypting it.
4. Clients can perform multiple local iterations on their data before sending updates to the central server. This can help ensure that sensitive support vector information is not exposed in the initial rounds of communication.
5. Instead of transmitting support vectors to the central server, clients can transmit only the parameters of the trained SVM model, such as the weight vectors.
6. Instead of sending raw support vectors, clients can use privacy-preserving data aggregation techniques like secure multi-party computation (SMPC) to collaboratively build the SVM model without revealing the individual support vectors.
7. Clients can preprocess their data locally to remove or obfuscate any sensitive information.

What is the key idea behind k-anonymity?

Ensuring that each record is indistinguishable from at least k-1 other records based on quasi-identifiers

What is the main limitation of k-anonymity?

It is vulnerable to attacks using background knowledge.

What does l-diversity aim to achieve?

Ensuring sufficient diversity of sensitive attribute values within each group.

What is the distinction between distinct l-diversity and entropy l-diversity?

Distinct l-diversity requires at least l distinct sensitive values, entropy l-diversity uses entropy to measure diversity.

What is the main limitation of l-diversity?

It does not provide a quantitative measure of privacy.

What is the basic idea of differential privacy?

Add calibrated noise to data to prevent identification of individuals.

What is a real-world application of differential privacy?

US Census Bureau uses it to publish anonymized data.

What is the “right to be forgotten” in the context of machine learning?

The ability to have your data deleted.

Differentiate between anonymity and k-anonymity in the context of data privacy.

Anonymity: identity completely hidden. K-anonymity: record indistinguishable from at least k others.

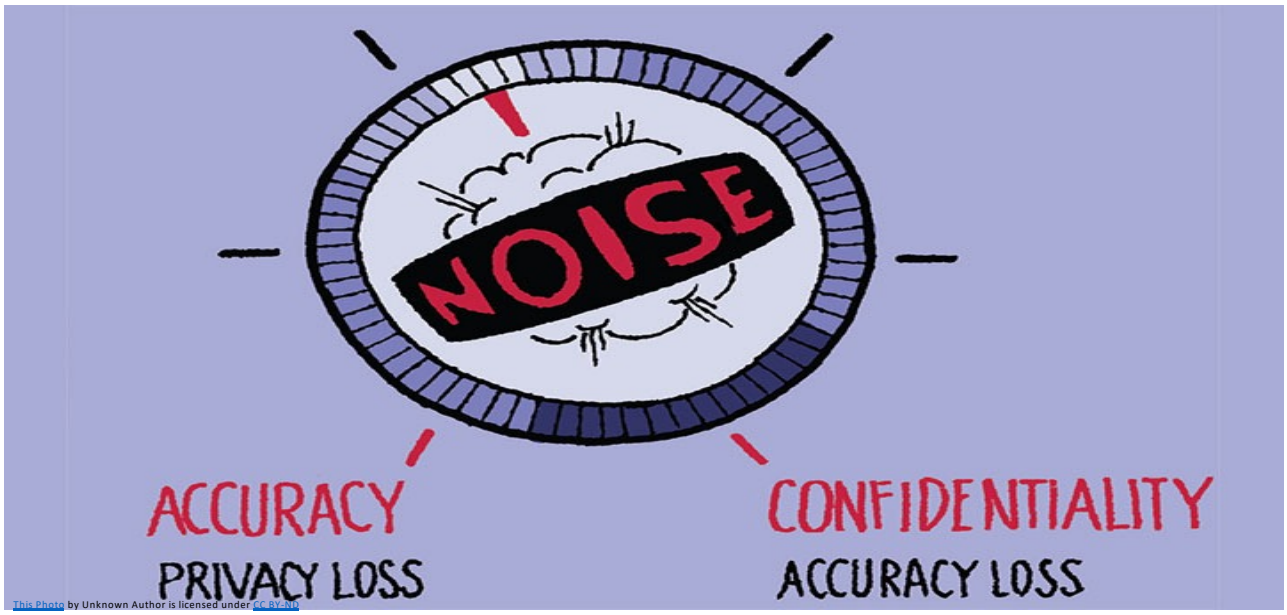
How does the L-diversity privacy model improve upon k-anonymity limitations?

Requires diverse sensitive attributes within anonymized groups, making re-identification harder.

Give an example of a homogeneity attack that could compromise anonymized data.

All records in a small anonymized group share a specific sensitive value, revealing information about individuals.

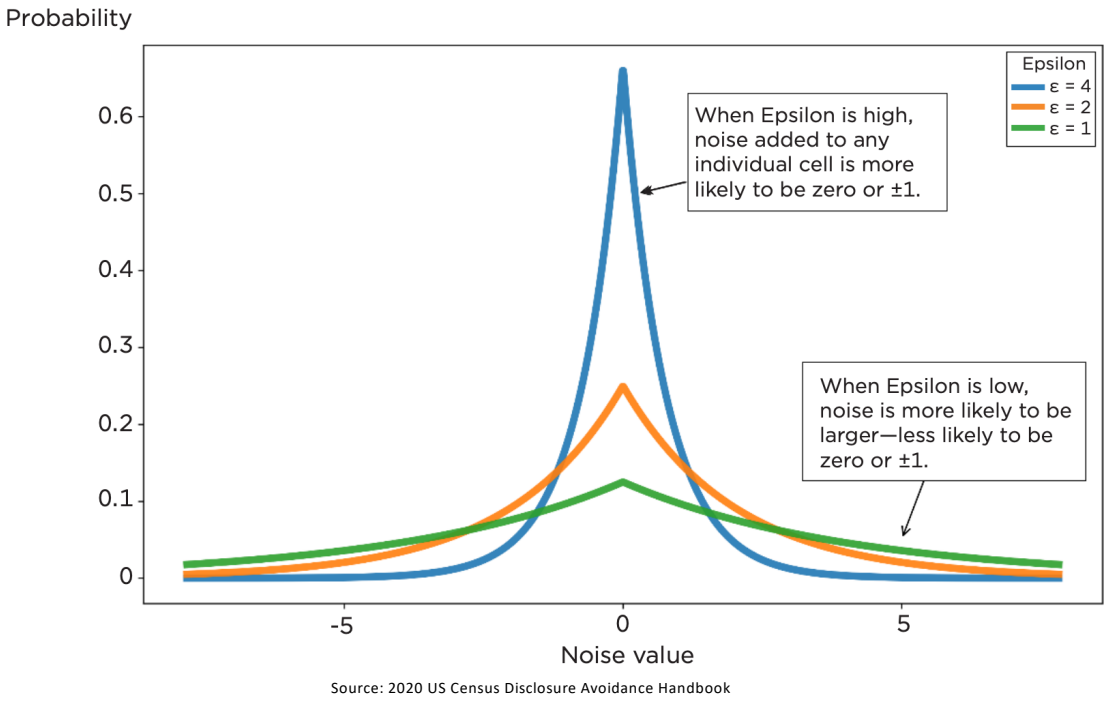
What is the Accuracy – Privacy Trade-off?



Laplace Distribution

$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

$$= \begin{cases} \frac{1}{2}\exp\left(\frac{x-\mu}{b}\right) & \text{if } x < \mu \\ 1 - \frac{1}{2}\exp\left(-\frac{x-\mu}{b}\right) & \text{if } x \geq \mu \end{cases}$$



Source: 2020 US Census Disclosure Avoidance Handbook

Adding Laplacian noise

Original Data:

ID	Age	Income
0	25	50000
1	32	72000
2	41	85000
3	28	61000

After adding Laplacian Noise:

ID	Age	Income
0	28.7	49522.0
1	30.2	73898.2
2	42.8	83327.3
3	25.3	60242.1

Code for adding Laplacian noise

- import pandas as pd
- import numpy as np
- data = {'Age': [25, 32, 41, 28], 'Income': [50000, 72000, 85000, 61000]}
- df = pd.DataFrame(data)
- # Define privacy parameters (epsilon for privacy and scale for noise)
- epsilon = 0.5
- scale = 2 / epsilon # Laplace scale
- # Generate Laplacian noise
- noise = np.random.laplace(loc=0, scale=scale, size=df.shape)
- # Add noise to each column
- df_noisy = df.copy()
- for col in df.select_dtypes(include=[np.number]):
- df_noisy[col] += noise[:, df.columns.get_loc(col)]

Drawbacks of differential privacy in machine learning?

Potential reduction in model accuracy and increased computational complexity.

Can differential privacy be combined with other techniques?

Yes, for layered protection (e.g., encryption and access control).

Why is the concept of neighboring data sets crucial?

Differential privacy guarantees similar algorithm outputs for datasets differing by only one record are similar.

Emerging trends in applying machine learning for sustainability?

Explainable AI (XAI) for understanding models, federated learning for privacy-preserving training, and reinforcement learning for complex resource management.

What is the meaning of the privacy parameter ε in differential privacy?

A tunable parameter that controls the trade-off between privacy and utility.

What is the sequential composition property of differential privacy?

The privacy guarantees degrade linearly when multiple differentially private mechanisms are applied.

What is the key idea behind machine unlearning?

Deliberately removing or updating knowledge or data from AI models and systems.

What is the main challenge in machine unlearning?

Non-deterministic training makes it difficult to understand the impact of select data items.

What is the goal of machine unlearning?

Achieving comparable accuracy, minimum time unlearning, provable guarantees, and model agnosticism.

What is the SISA (Sharded, Isolated, Sliced, and Aggregated) approach for machine unlearning?

A technique that shards and isolates data to enable faster unlearning.

What is the purpose of the "Right to be Forgotten" in the context of machine learning?

Enabling individuals to request the removal of their personal information from AI models.

What is the significance of the homogeneity attack on k-anonymity?

It demonstrates that k-anonymity can leak information due to lack of diversity in sensitive attributes.

What is the challenge in achieving machine unlearning through retraining?

Retraining the model from scratch after removing the data is computationally expensive.

What is the significance of the example involving the Strava fitness app data leak?

It demonstrates the potential privacy risks of unintentionally sharing sensitive location data.