AINA

AI and Analytics

Volume 1 • Edition 2019-20

Annual analytics magazine from the students of PGDBA

Post Graduate Diploma in Business Analytics: *Jointly offered by IIM Calcutta, ISI Kolkata, IIT Kharagpur*

PRIVACY IN AN OPEN WORLD

BY CHANDU V. GRANDHI

User info such as names, gender, zip codes, IP addresses etc. are collected, analyzed & stored alongside similar confidential information by many firms. In most cases, cryptography is employed for data encryption to make it unintelligible without the decryption key. Such data can lead to serious problems when in wrong hands. Hence, "Data Privacy" is quintessential today.

Data privacy and data security are often confused to be synonymous. While data privacy deals with the proper usage, collection, retention, deletion, and storage of data, data security is all about policies, method and means to secure personal data.

With increasing computational ability, reverse engineering or decoding is a real threat. Thanks to the legal developments around the data privacy, regulations have been constantly upgrading to match today's standards. The European Commission drafted the General Data Privacy Regulation or GDPR to identify & take appropriate measures against those responsible in such cases of breach or mishandling of data. It is mandatory for all organizations established or operating in EU to abide by these.

Facebook-Cambridge Analytica scandal which involved psychological profiling during the US elections'16 is largely credited with raising several privacy concerns. This sparked a new debate about the competence of firms against policy violations.

The concept of "K-anonymity" in privacy protection is quite popular. As per common definition: "If the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose info also appear in the release, then k-anonymity is achieved."

"Differential privacy" enable sharing of data while withholding the sensitive information of individuals in dataset. This results in the same output both with the presence or absence of the user information in the model & guarantees protection against such attacks measured using "privacy loss" metric. It is important to understand that differential privacy models are used for the protection of private data only.

Several edge-computing techniques have been deployed to store and process data at end-user devices. Apple developed strong computational capabilities and development libraries on the A13 bionic chip, thereby eliminating the need to upload data to cloud for processing.

As Google phrased it in their 2017 paper, "Federated learning" (FL) is a distributed machine learning approach which enables model training on a large corpus of decentralized data residing on end devices such as mobile phones. These companies assimilate galactic amounts of data whose centralization is usually privacy intrusive. FL facilitates data distribution globally, avoiding the collection at a single source. Information from the "teacher" models is processed and the outputs are used to train a "student" model that receives the inputs from these teachers. Well designed FL methods can ensure utmost privacy even against generative means of reverse engineering. With people often sharing crucial data such as location and contacts to several mobile applications with least deliberation, it is often the individual's role to selectively secure their data to ensure privacy, especially in the post-privacy age of hyper-customization & exceptionally accurate recommendation engines.