



## ABSTRACT

The increased smartphone penetration creates an abundant amount of near-real-time human behavior smartphone dataset. Nevertheless, studies show that this dataset is not entirely secure. The leakage in the smartphone dataset is alarming; since this data holds much information about users that is able to, for instance, identify specific users among crowd in user identification attempts.

The concept of user identification becomes vital, since the main idea of privacy is ‘the right to be let alone’. Yet various studies has proven that smartphone data, e.g., smartphone usage data, sensor data, and user input data can identify users. Nevertheless, a user identification study exploring smartphone activity data is not yet available. As the recent data leakage in Google Takeout service raises questions in the security of the data hold in third parties’ services, we intend to show that anonymous smartphone activity data stored in these external services can violate privacy by linking the data back to an individual.

We obtained a total accumulation 551 days of dataset from seven users using App Usage apps on Android platform. This dataset is calculated to generate User Profiles, which act as the users’ fingerprint or identity that differentiate one user from others. Using statistic approach, Convolution Neural Network (CNN) algorithms, and k-Nearest Neighbour (kNN) algorithms, we obtain average accuracy of 87-89% by observing the percentage usage duration of five most-used applications. These approaches easily identify five from seven users that have absolute distinguishable features, while while the remain users are slightly similar and required further observation in future works.

We have successfully proven that smartphone activity data can firmly identify users. Thus, we hope that the result of this study can then intrigue new discussion regarding smartphone data privacy.

**Keywords:** Smartphone Activity Data, User Identification, Privacy, CNN, Statistics, kNN



## INTISARI

Jejak data digital yang berlalu lalang di dunia maya sangat masif. Akan tetapi, studi membuktikan bahwa data tersebut tidaklah sepenuhnya terlindungi. Berbagai kasus kebocoran data terus muncul dan menimbulkan kewaspadaan, karena dengan mengolah data tersebut, beragam masalah privasi dapat muncul – contohnya dengan melakukan identifikasi pengguna.

Konsep identifikasi pengguna menjadi masalah karena salah satu ide utama dari privasi adalah hak untuk dibiarkan sendiri (*‘the right to be let alone’*). Sehingga, informasi yang tersebar tentang seseorang tidak seharusnya dapat mengidentifikasi orang tersebut. Beragam studi telah membuktikan bahwa data dari gawai dapat melakukan identifikasi pengguna dengan menggunakan data penggunaan gawai, data dari sensor, maupun data dari masukan pengguna. Namun, belum terdapat satupun penelitian yang menelaah data dari aktivitas penggunaan gawai. Sedangkan data aktivitas penggunaan gawai tersebut disimpan oleh pihak ketiga seperti Google – yang pada Februari 2020, Google memberikan notifikasi bahwa data pengguna yang disimpan pada layanan Google Takeout besar kemungkinan tersebar ke orang asing. Sehingga, penelitian ini bermaksud menunjukkan bahwa data aktivitas penggunaan gawai, walaupun disimpan dengan skema anonim, tetap dapat mengidentifikasi individu spesifik dengan menggunakan metode identifikasi pengguna.

Penelitian ini mengumpulkan total akumulasi 551 hari dataset dari tujuh pengguna menggunakan aplikasi *App Usage* pada platform Android. Dataset tersebut lalu diolah untuk membuat *User Profiles* yang berperan sebagai sidik jari yang membedakan satu pengguna dengan pengguna lain. Menggunakan perhitungan statistik, *Convolution Neural Network* (CNN), dan *k-Nearest Neighbour* (kNN), penelitian ini mendapatkan rata-rata akurasi senilai 87-89 persen dengan mengobservasi aktivitas penggunaan pada lima aplikasi utama. Lima dari tujuh pengguna terbukti dapat teridentifikasi dengan jelas, dan dua pengguna lainnya mempunyai karakteristik yang mirip sehingga membutuhkan observasi lebih lanjut di penelitian selanjutnya.

Penelitian ini membuktikan bahwa aktivitas penggunaan gawai dapat dengan jelas melakukan identifikasi pengguna. Sehingga, hasil dari penelitian ini diharapkan



**Identifikasi Pengguna Menggunakan Data Aktivitas Penggunaan Gawai**

SYAFIRA FITRI AULIYA, Ir. Lukito Edi Nugroho, M.Sc. Ph.D; Noor Akhmad Setiawan, S.T., M.T., Ph.D.

Universitas Gadjah Mada, 2021 | Diunduh dari <http://etd.repository.ugm.ac.id/>

UNIVERSITAS  
GADJAH MADA

dapat membuka diskusi baru mengenai keamanan dan privasi data gawai.

**Keywords:** Aktivitas Penggunaan Gawai, Identifikasi Pengguna, Privasi, CNN, Statistik, kNN