

V této práci navrhujeme klasifikační algoritmus založený na metodách strojového učení pro aplikace na populární operační systém Android, který má za cíl rozlišovat škodlivé aplikace od nezávadných. Extrakce příznaků pro strojové učení je založena na statické analýze bajtkódu se zaměřením na API volání a volání metod. Ukazujeme různé přístupy jak z nejčastějších API volání vytvořit seznam číselných příznaků. Dále zkoumáme specifika extrahovaných příznaků a jejich důležitost. Dataset, který je použit pro experimenty v této práci, obsahuje přes 200 000 vzorků, z nichž přibližně polovina je škodlivá a polovina nezávadná. Zkoušíme několik různých algoritmů strojového učení a vyhodnocujeme jejich kvalitu. Velikost našeho datasetu snižuje poměr šumu a poskytuje tak dobrý základ pro trénování klasifikačních modelů. Výsledky experimentů ukazují, že navrhovaný algoritmus má poměru vzorků, které byly chybně označeny jako škodlivé, pod 2,9 % přičemž správně nalezne přes 93,6 % malwaru.