# Abstract

Rozprawa przedstawia serię publikacji podejmujących wyzwanie konstruowania optymalnych globalnych modeli klasyfikacji na podstawie rozproszonych i heterogenicznych zbiorów danych, bez konieczności scentralizowanego dostępu do nich. W tym ujęciu dane rozproszone odnoszą się do zbiorów przechowywanych w wielu niezależnych źródłach, których nie można fizycznie skonsolidować, natomiast dane heterogeniczne to zbiory różniące się przestrzeniami cech oraz zbiorami obiektów w zależności od lokalizacji. Tradycyjne podejścia stosowane w rozproszonych środowiskach danych zakładają jednorodne przestrzenie cech oraz możliwość bezpośredniej agregacji danych, co czyni je niewystarczającymi w rzeczywistych scenariuszach obejmujących ograniczenia prywatności oraz strukturalna zmienność danych. Aby przezwyciężyć te ograniczenia, w rozprawie przedstawiono i przeanalizowano trzy komplementarne podejścia oparte na sieciach neuronowych, z których każde zostało zaprojektowane z myślą o wyzwaniach wynikających z niejednorodności i rozproszenia danych. Pierwsze podejście polega na integrowaniu predykcji uzyskiwanych z lokalnie trenowanych klasyfikatorów k-najbliższych sąsiadów poprzez sieć neuronową odpowiedzialną za fuzję niezależnych wyników. Pozwala to uzyskać spójną globalną klasyfikację przy jednoczesnym zachowaniu prywatności danych. Drugie podejście wykorzystuje mechanizm imputacji obiektów syntetycznych w celu harmonizacji różnic strukturalnych między lokalnymi zbiorami danych. Dzięki temu możliwa jest agregacja wag lokalnie wytrenowanych sieci w jeden wspólny model globalny. Trzecie podejście opiera się na technikach ekstrakcji cech, takich jak Parincipal Component Analysis (PCA), Singular Value Decomposition (SVD) oraz Uniform Manifold Approximation and Projection (UMAP), które pozwalają rzutować lokalne dane na ujednoliconą przestrzeń cech. Otrzymane modele są następnie integrowane za pomocą głosowania. Przeprowadzone eksperymenty pokazują, że przedstawione podejścia konsekwentnie zwiększają dokładność i odporność modeli klasyfikacji w porównaniu z klasycznymi metodami zespołowymi i metodami scentralizowanymi, zapewniając jednocześnie zachowanie prywatności danych oraz efektywność obliczeniową. Wyniki te wnoszą nową wiedzę do obszaru rozproszonego uczenia opartego na sieciach neuronowych i stanowią podstawę skalowalnych, zorientowanych na ochronę prywatności metod integracji danych w środowiskach rozproszonych.