Miniprojekt 3

Plik `phishing.data` zawiera dane o stronach internetowych wraz z informacją, czy dana strona służy do phishingu (1), czy też jest to strona bezpieczna (-1). Każdy wiersz zawiera 31 liczb. Pierwsze 30 z nich to cechy, spośród których:

- pierwsze 21 przyjmuje wartości ze zbioru $\{-1, 1\}$,
- kolejnych 8 cech przyjmuje wartości ze zbioru $\{-1, 0, 1\}$,
- ostatnia cecha przyjmuje wartość ze zbioru $\{0, 1\}$.

W ostatniej kolumnie znajduje się informacja, czy dana strona jest stroną phishingową.

W oparciu o powyższe dane napisz trzy programy uczące, które zwrócą jak najlepsze klasyfikatory stron podejrzanych o phishing. Jeden z programów powinien bazować na **metodzie wektorów nośnych** (SVM), przy czym można korzystać z dowolnych funkcji jądrowych. Pozostałe dwa programy powinny być oparte na dwóch spośród poniższych **klasyfikatorów nieliniowych**:

- metoda k najbliższych sąsiadów (kNN),
- drzewa decyzyjne,
- lasy losowe,
- AdaBoost,
- sieć neuronowa.

Wykorzystaj część zbioru obserwacji jako zbiór treningowy, odpowiednio mniejszą część jako zbiór walidacyjny, a pozostałą część jako zbiór testowy.

W raporcie opisz zastosowane algorytmy uzasadniając wybór parametrów modeli. Dokonaj analizy uzyskanych wyników oraz przedstaw je na odpowiednich wykresach.

W ramach ciekawostki zapoznaj się z dołączonymi artykułami. Czy uzyskane przez Ciebie wyniki są zbliżone do wyników przedstawionych w którymś z artykułów?