

Student: Mateusz Kuchta

Kierunek studiów: Big Data

Tryb studiów: niestacjonarne

Planowany termin obrony: czerwiec 2023

Proponowany tytuł pracy:

„Wpływ wrogiego uczenia maszynowego na modele estymacji wiarygodności kredytowej”

Celem pracy jest budowa modelu predykcyjnego do wykorzystania w dziedzinie Credit Scoring'u. Model zostanie opracowany przy wykorzystaniu algorytmu drzewa decyzyjnego. Odpowiednio nauczony model zostanie następnie poddany „atakowi”, znanemu jako „wrogie uczenie maszynowe”. Wynikiem pracy powinny być wnioski dotyczące wpływu tego typu „oszustwa” na model w zależności od skali ataku.

W pierwszym rozdziale zostanie opisany temat wiarygodności kredytowej, używanych do jej oszacowania modeli, a także przyczyny powstania problemu wrogiego uczenia maszynowego i dotychczasowe dokonania dotyczące badań w tej dziedzinie. Obejmie on szeroki przegląd literatury tj. badań naukowych i publikacji, jak także przywoła źródła występujące w formie video udostępnione przez np. Uniwersytet Stanford.

W drugim rozdziale zostanie przedstawiony bardziej konkretny i szczegółowy zakres badania, co obejmuje wybranie zbioru danych i uzasadnienie jego wyboru, selekcja technik, które zostaną wykorzystane do przetworzenia i analizy danych wejściowych do modelu, a także plan przetestowania modelu pod kątem wrażliwości na wrogie ataki. Postawione zostaną hipotezy badawcze wraz z uzasadnieniem potencjalnej wartości płynącej z ich weryfikacji.

Kolejny rozdział będzie stanowił opis wykorzystania narzędzi do celu budowy modelu drzewa decyzyjnego. Do jego realizacji zostanie wykorzystany Jupyter Notebook, a jako język programowania zastosowany będzie Python. Na tym etapie zostanie dokonana eksploracyjna analiza danych oraz zbudowany będzie model predykcyjny. Rozdział zwieńczy podsumowanie otrzymanego modelu.

Ostatni rozdział zostanie rozpoczęty od zaplanowania „ataków” na stworzony model. Poprzez wprowadzenie fałszywych danych zostanie zbadana jego odporność na ingerencje o różnym charakterze, co również zostanie zaimplementowane w języku Python. Celem testów będzie znalezienie wrażliwych punktów utworzonego modelu i ostateczna weryfikacja postawionych hipotez badawczych.

Do wykonania pracy zostanie wykorzystane również oprogramowanie Overleaf, celem opanowania przez studenta nowego narzędzia do edycji tekstu, a także inteligentnego zarządzania bibliografią, zachowując przy tym wysoki standard jakości dokumentu.

Ramowy spis treści

Wstęp

1. Część teoretyczna
 - 1.1. Credit Scoring
 - 1.2. Uczenie Maszynowe
 - 1.3. Adversarial Machine Learning – Wrogie Uczenie Maszynowe
2. Wprowadzenie do budowy modelu drzewa decyzyjnego w Credit Scoring
 - 2.1. Osiągnięcia w dziedzinie wykorzystania drzew decyzyjnych w Credit Scoring
 - 2.2. Wybór i opis wykorzystanego zbioru danych
 - 2.3. Koncepcja przeprowadzenia części praktycznej
3. Budowa modelu drzewa decyzyjnego do celu Credit Scoring
 - 3.1. Wykorzystane narzędzia i technologie
 - 3.2. Budowa modelu
 - 3.3. Analiza wyników
4. Atak na opracowany model
 - 4.1. Strategia badania odporności modelu
 - 4.2. Implementacja wybranych technik ataku
 - 4.3. Analiza wyników i weryfikacja hipotez badawczych

Wnioski

Bibliografia

Spis rysunków

Spis tabel

Załączniki