Studium Magisterskie

Kierunek: Big Data – Analiza Danych

Mateusz Kuchta

Nr 116111

„Wpływ wrogiego uczenia maszynowego na modele estymacji wiarygodności kredytowej”

Praca magisterska
pod kierunkiem naukowym
dr Mariusza Rafało w Kolegium Analiz Ekonomicznych SGH

Warszawa 2023

Spis treści

Wstęp

[2 STRONY WSTĘPU]

1 Część teoretyczna

[WSTĘP DO ROZDZIAŁU]

1.1 Credit Scoring

Ocena wiarygodności kredytowej jest stosowana na całym świecie do przetwarzania wielu rodzajów pożyczek. Jest ona wykorzystywana najszerzej i z największym powodzeniem w przypadku osobistych kart kredytowych oraz kredytów hipotecznych. Ryzyko spłaty tych produktów jest ściśle powiązane z weryfikowalnymi czynnikami, takimi jak dochód, informacje biura kredytowego i czynniki demograficzne, takie jak wiek, wykształcenie i status cywilny (Dean Caire, 2006). Nieraz trudno jest ocenić, czy dana osoba zasługuje na zaufanie, czy może tym razem bank powinien się wstrzymać, nie ryzykując problemami ze spłatą danego kredytobiorcy, zarazem rezygnując z potencjalnego zysku. Dokładne określenie granicy między dobrym, a złym klientem jest trudne dla nawet najbardziej doświadczonych pracowników instytucji finansowych. Zwiększona konkurencja i rosnąca presja na generowanie przychodów skłoniły instytucje udzielające kredytów do poszukiwania skutecznych sposobów pozyskiwania nowych klientów o dobrej zdolności kredytowej, a przy tym jednocześnie kontroli zysków i strat. Agresywne działania marketingowe przyniosły efekt głębszej penetracji danych potencjalnych klientów, a potrzeba szybkiego i efektywnego ich przetwarzania doprowadziła do rosnącej automatyzacji procesu składania wniosków kredytowych i ubezpieczeniowych oraz ich rozpatrywania (Siddiqi, 2016).

1.1.1 Credit Scoring – definicja i cechy

Credit Scoring można zdefiniować jako zespół statystycznych metod, które są używane w celu przewidywania prawdopodobieństwa, iż wnioskodawca kredytu nie wywiąże się w terminie z zaciągniętych zobowiązań. Pomaga to ustalić, czy kredyt powinien być przyznany kredytobiorcy. Scoring kredytowy jest jedną z metod systematycznej oceny, która została uznana za istotnie wpływającą na obniżenie poziomu ryzyka kredytowego w banku. W literaturze można znaleźć wiele różnych definicji scoringu. Wynika to z faktu, że jest to pojęcie w znacznym stopniu subiektywne, a banki mają dużą swobodę w stosowaniu technik punktowych. Ważne jest, aby model scoringowy charakteryzował się wysoką skutecznością w oddzielaniu klientów dobrych od złych (Wysiński, 2013). Jest to zatem kluczowe narzędzie w rękach instytucji finansowych dające możliwość ograniczania potencjalnego ryzyka

współpracy z niewiarygodnym klientem, przy jednoczesnym osiąganiu maksymalnych zysków zawierając umowy z wartościowymi kredytobiorcami (direct.money.pl, 2022).

Credit Scoring charakteryzuje się kilkoma podstawowymi cechami (Encyklopedia Zarządzania, 2020):

- Dane historyczne jako baza – porównuje się ze sobą charakterystyki grup rzetelnych kredytobiorców z nierzetelnymi i na tej podstawie dokonuje się oceny, czy potencjalny klient będzie terminowo spłacał zaciągnięte zobowiązanie, czy też może istnieje wysokie prawdopodobieństwo, że zachowa się on podobnie jak usługobiorcy mający problemy z uiszczaniem kolejnych rat na czas;
- Okresowość – mechanizmy wyliczania oceny wymagają częstych aktualizacji o nowe dane, w celu zapewnienia jak najwyższej skuteczności otrzymanego wyniku;
- Rzetelne zbadanie zdolności kredytowej kredytobiorców jako środek do celu jakim jest ochrona interesów kredytodawcy;
- Realizowany za pomocą zaakceptowanych i zatwierdzonych metod statystycznych.

Punktowa ocena wiarygodności kredytowej budowana jest na zasadzie przyznawania punktów za określone cechy kredytobiorcy, gdzie im wyższy wynik wnioskodawca osiągnie, tym większa szansa, że spłaci kredyt w terminie. Tabela punktowa tworzona jest na podstawie analizy statystycznej bazy danych klientów z przeszłości, gdzie poszukuje się cech, które w jak najlepszy sposób oddzielają od siebie dobrych i złych biorców kredytowych (bankier.pl, 2012).

Dla pewnego modelu, wpływ posiadania własnego mieszkania na spłacenie zobowiązania bez opóźnień zwizualizowano na Rysunek 1:

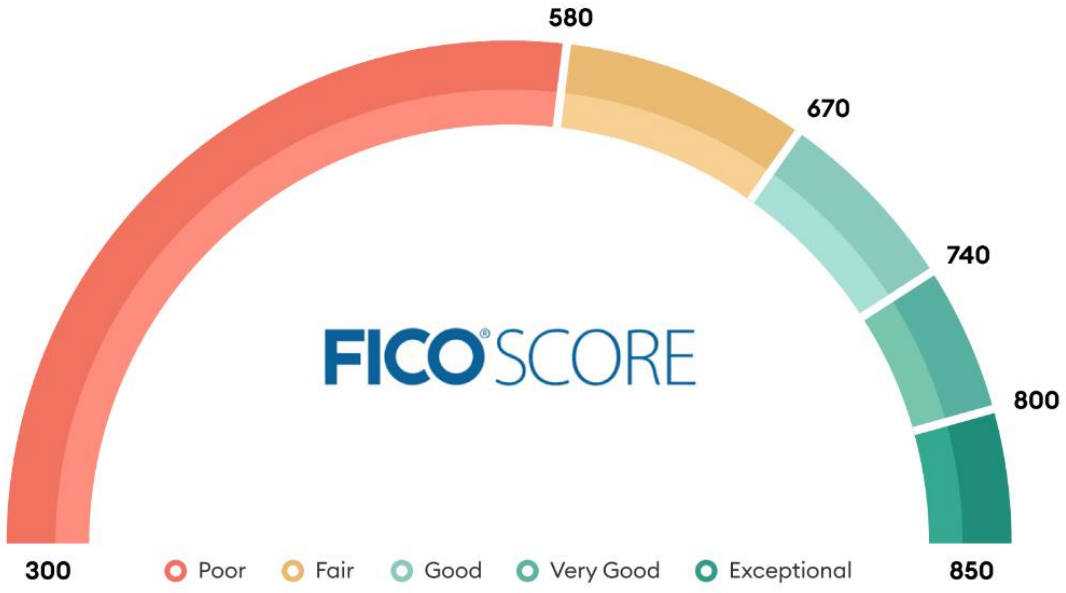
Rysunek 1. Przykład cechy wykorzystanej w Credit Scoring'u *(bankier.pl, 2012)*

Na Rysunek **1** widać, że wśród osób mających problemy ze spłatą pożyczki aż 95% stanowią osoby wynajmujące nieruchomość. Zatem posiadanie własnego mieszkania będzie stawiać w uprzywilejowanej pozycji potencjalnych klientów banku i otrzymają oni wyższą ocenę za tę cechę w ogólnym scoring'u. Taką logiką kierują się zarówno banki, jak i BIK – Biuro Informacji Kredytowej. Sposób wyliczania oceny punktowej często jest tajemnicą w instytucjach finansowych i zwykle nie dowiemy się jaki score otrzymaliśmy oraz co jest powodem takiego wyniku. Jednakże w Biurze Informacji Kredytowej, mimo iż nie poznamy pełnej logiki oceniania, pośrednio wiadome są najważniejsze kryteria oszacowań. Wśród najważniejszych z nich należy wymienić (bankier.pl, 2012):

- Liczba nieterminowo spłacanych zobowiązań;
- Wysokość zobowiązań (np. na karcie kredytowej lub debetowej);
- Czas jaki zwlekamy ze spłatą zobowiązania;
- Czas jaki upłynął od ostatniego wykroczenia.

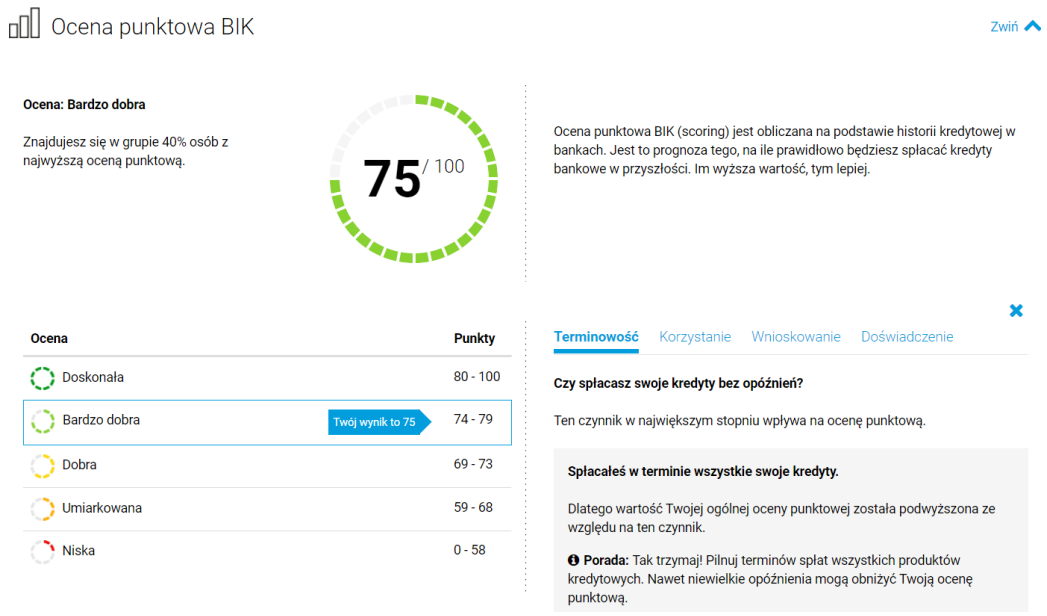
1.1.2 Wyznaczanie oceny punktowej

W Stanach Zjednoczonych score kredytowy zawiera się w granicach od 300 do 850 punktów, gdzie w zależności od instytucji za odpowiednio dobry wynik traktuje się wartości powyżej 661-670 punktów (experian.com, 2021).



Rysunek 2. Diagram oceny wiarygodności kredytowej według firmy FICO *(forbes.com, 2021)*

Od 21 grudnia 2017 roku BIK generuje ocenę punktową w nieco inny sposób. Dotychczas score zawierał się w zakresie od 192 do 631 punktów, co wizualizowane było za pomocą gwiazdek, gdzie im więcej gwiazdek, tym wyższa szansa na otrzymanie kredytu. Aby uczynić reprezentację graficzną bardziej czytelną dla osób fizycznych, ocenę przekształca się do postaci od 1 do 100 (totalmoney.pl, 2020). Zostało to zilustrowane jest w postaci wykresu kołowego na Rysunek 3:



Rysunek 3. Wizualizacja oceny punktowej w BIK (źródło opracowanie własne)

Przed zmianą z grudnia 2017, Biuro Informacji Kredytowej do wyliczania oceny punktowej wykorzystywało tzw. Model II generacji o nazwie BIKSco CreditRisk. Dla tego algorytmu zakres możliwych do uzyskania rezultatów zawierał się pomiędzy 192, a 631 punktów i nie był w żaden sposób przekształcany. W najnowszych raportach wykorzystywany jest model III generacji o nazwie BIKSco CreditRisk 3, a punktacja zawiera się od 98 do 711 punktów. Jednakże otrzymany wynik ulega przekształceniu i tak jak jest to widoczne na Rysunek 3, przedstawiony jest w zakresie od 1 do 100, co może mylnie wskazywać, że jest to procent maksymalnej, możliwej do uzyskania punktacji (scoringexpert.pl, 2018).

Bazując na cytowanym źródle, wynik wyliczonej punktacji podlega procesowi normalizacji według Równanie 1:

Równanie 1. Wzór na przekształcenie oceny wyliczonej z modelu BIKSco CreditRisk 3 do przedziału znormalizowanego 1-100 *(experian.com, 2021)*

$$S_{new} = \frac{S - \min}{max - min} * (new_{max} - new_{min}) + new_{min}$$

gdzie:

S_{new} – ocena punktowa z raportu BIK;

S – oryginalna ocena punktowa z modelu BIKSco CreditRisk 3;

min – minimalna wartość punktów z modelu BIKSco CreditRisk 3 (wynosi 98);

max – maksymalna wartość punktów z modelu BIKSco CreditRisk 3 (wynosi 711);

new_{max} – maksymalna wartość punktów z nowego zakresu (wynosi 100);

new_{min} – minimalna wartość punktów z nowego zakresu (wynosi 1).

Na podstawie wartości widocznej dla osoby fizycznej (punktacji z przedziału 1-100), można uzyskać wynik wyliczony z algorytmu na podstawie Równanie 2:

Równanie 2. Wzór na wyliczenie wartości otrzymanej z modelu BIKSco CreditRisk 3 na podstawie wartości uzyskanej z Biura Informacji Kredytowej *(experian.com, 2021)*

$$S = \frac{613 * S_{New} + 9089}{99}$$

Biuro Informacji Kredytowej proponuje swoją interpretację znormalizowanej oceny punktowej, przedstawioną na Rysunek 3. Wizualizacja oceny punktowej w BIK (źródło opracowanie własne), według której potencjalny kredytobiorcę może ocenić swoje aktualne szanse na otrzymanie pożyczki.

Na podstawie cytowanej Tabela 1, można zinterpretować również ocenę nieznormalizowaną:

Tabela 1. Interpretacja oceny punktowej BIK *(experian.com, 2021)*

Ocena punktowa BIK	**Słowna ocena**	**Komentarz**
550+ (74+)	Bardzo dobra	Twój „scoring BIK" przewyższa średni „scoring BIK" Polaków. W ocenie banków Twoja wiarygodność kredytowa powinna być bardzo wysoka
500-549 (66-73)	Dobra	Twój „scoring BIK" oscyluje blisko średniego „scoringu BIK" Polaków. Banki zapewne ocenią Cię jako rzetelnego kredytobiorcę.
400-499 (50-65)	Przeciętna	Twój „scoring BIK" jest poniżej średniego „scoringu BIK" Polaków. Tylko część banków oceni Twoją wiarygodność kredytową jako wystarczającą do uzyskania kredytu
<400 (<50)	Słaba	Twój „scoring BIK" jest znacznie poniżej średniego „scoringu BIK" Polaków. Taki scoring wskazuje, że jesteś ryzykownym kredytobiorcą dla banków. Możesz więc mieć problem z uzyskaniem kredytu, jeśli bank oprze się na „scoringu BIK" przy ocenie Twojego ryzyka kredytowego

Za jedną z najbardziej znanych metod Credit Scoring'u jest przytaczana na łamach tej pracy amerykańska metoda FICO (Fair, Isaac and Company). Zdefiniowana w 1989 roku, opiera się na 5 czynnikach (pl.economy-pedia.com, 2021):

- Historia płatności (35% punktów) – na bazie dotychczasowych zobowiązań ocenia się, czy dana osoba wywiązuje się z nich na czas;
- Wykorzystanie kredytu (30%) – jeśli potencjalny klient instytucji finansowej dotychczas wykorzystywał niewielki procent dostępnych limitów kredytowych (np. limit na karcie kredytowej), ma on większe szanse na wyższą ocenę;

- Długość historii kredytowej (15%) – jeśli wnioskodawca jest doświadczonym pożyczkobiorcą i przez długi czas poprawnie wywiązuje się z zobowiązań otrzymuje wyższą notę w tabeli punktowej;
- Nowe kredyty (10%) – złożenie przez wielu wniosków kredytowych przez osobę poszukującą kredytu, może wzbudzać wątpliwości instytucji przed udzieleniem pożyczki;
- Rodzaje wykorzystanego kredytu (10%) – dla banków mile widziane jest doświadczenie wnioskującego w zarządzaniu różnymi rodzajami kredytów (karta kredytowa, kredyt hipoteczny, pożyczka gotówkowa itp.).

Najbardziej znane techniki stosowane do opracowania kart scoringowych to dyskryminacja statystyczna i metody klasyfikacji. Należą do nich modele regresji liniowej (łatwo interpretowalne, oparte na metodzie minimalizacji sumy kwadratów reszt), analiza dyskryminacyjna (odmiana regresji stosowana do klasyfikacji), modele logitowe i probitowe (maksymalizacja prawdopodobieństwa zaobserwowania wartości), oraz tzw. modele eksperckie (np. proces analizy hierarchicznej – AHP) (The World Bank Group, 2019).

Rozwój Credit Scoringu jest jednym z powodów, dla których rynek kredytów konsumenckich w Stanach Zjednoczonych w latach 90. XX w. eksplodował. Kredytodawcy czuli się bardziej pewni udzielania pożyczek szerszym grupom ludzi, ponieważ mieli dokładniejsze narzędzie do pomiaru ryzyka. Scoring kredytowy pozwolił im również na szybsze podejmowanie decyzji, umożliwiając im rozpatrzenie większej liczby wniosków, czego rezultatem był bezprecedensowy wzrost ilości dostępnych kredytów konsumenckich (Weston, 2012). Banki bardzo skrupulatnie podchodzą do operowania swoimi pieniędzmi, dokładnie „prześwietlając" swoich klientów pod kątem wypłacalności. Polskie instytucje finansowe często posiłkują się opinią BIK, jednakże równie chętnie stosują również swoje algorytmy oceny, których szczegółów zwykle szerzej nie udostępniają. Kolejnym krokiem pracy było dokładne zapoznanie się z rozwojem wykorzystania narzędzi Big Data w dziedzinie Credit Scoring.

1.1.3 Score kredytowy w erze Big Data

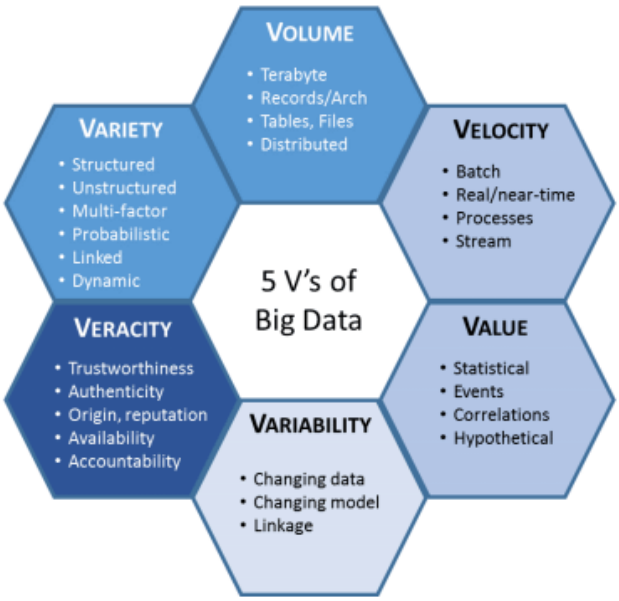
Dziedzina Credit Scoringu nie należy do odkryć bieżącej dekady, a jej początki można upatrywać w połowie XX wieku wraz z utworzeniem w 1956 roku wspominanej już wcześniej firmy FICO. Przedsiębiorstwo założyli inżynier Bill Fair oraz matematyk Earl Judson Isaac (Wikipedia, 2022). Firma zajmowała się budową systemów scoringowych, jednakże przez

długi czas stosowano zapis papierowy (StatSoft, 2010). Początkowo zajmowano się głównie procesami weryfikacji wniosków kredytowych w bankach, stosując proste tabele punktowe i głównie opierając się na tzw. metodzie eksperckiej (Thonabauer G., 2004). Sposób musiał być na tyle łatwy i intuicyjny, aby dawać możliwość obiektywnej oceny zdolności potencjalnego kredytobiorcy do wywiązania się ze zobowiązania kredytowego również mniej doświadczonym i wykwalifikowanym pracownikom banku (Thomas L.C., 2002).

Wraz z rozwojem technik informatycznych rozpoczęto wdrażanie bardziej zautomatyzowanych procesów scoringowych. Najpowszechniej do tego celu wykorzystywano model regresji logistycznej, jednakże dziś stosuje się szeroką gamę różnych metod predykcyjnych tj. sieci neuronowe, lasy losowe czy drzewa decyzyjne. Taka transformacja może dawać wrażenie, iż Credit Scoring usuwa się w cień na korzyść Big Data, choć w rzeczywistości staje się jedną z poddziedzin szerokiej tematyki „dużych danych" (Przanowski, 2014). Lecz czym właściwie jest Big Data?

Pojęcie to powinno iść w parze z rozbudowanymi systemami informatycznymi, które dają możliwość przetwarzania dużych danych. Często spotykanym jest tzw. 5V Big Data (zwizualizowane na Rysunek 4), dające ogólne spojrzenie na podstawowe cechy tejże dziedziny (techtarget.com, 2021):

- Volume (ang. wolumen, wielkość) – jeśli mamy do czynienia z subiektywnie dużą ilością danych (co dokładnie oznacza „duża ilość" nie zostało precyzyjnie zdefiniowane), możemy nazwać ich przetwarzanie jako Big Data;
- Velocity (ang. szybkość) – ze względu na ich ilość, dane muszą być odpowiednio szybko procesowane, najczęściej w czasie rzeczywistym;
- Variety (ang. różnorodność) – technologia musi radzić sobie z operowaniem na danych zróżnicowanego, nie koniecznie uporządkowanego typu;
- Veracity (ang. wiarygodność) – dane najczęściej nie są wysokiej jakości, a ich braki czy też błędne informacje w nich zawarte stawiają wymóg odpowiedniej odporności na tego rodzaju zaburzenia;
- Value (ang. wartość) – kluczowa przy przetwarzaniu danych jest możliwość uzyskania na ich bazie istotnych informacji, które mogą wnieść pewną wartość dla przedsiębiorstwa, dokonującemu lub zlecającemu wykonanie takich analiz.



Rysunek 4. Schemat 5V Big Data *(researchgate.net, 2017)*

Początkowo Credit Scoring specjalizował się głównie we wspomaganiu procesów decyzyjnych w bankach, a narzędzia Big Data stosowane były w globalnych firmach świadczących usługi w świecie wirtualnym tj. Google, Amazon czy Facebook. Z kolei w Polsce zarządzaniem dużymi danymi na poważnie zainteresowały się jako pierwsze Onet czy portal Nasza Klasa (Przanowski, 2014). Mimo zainteresowania różnymi branżami, Big Data i Credit Scoring poruszają podobne problemy ze strony merytorycznej, gdzie głównym i najpoważniejszym problemem zawsze był kluczowy element ich funkcjonowania – dane.

Modele Credit Scoring służą do prognozowania zjawisk na podstawie dotychczas zaobserwowanej i zebranej historii danych. Proces spłacania kredytów najczęściej trwa wiele lat, zatem potrzeba dużo czasu, aby zebrać dostatecznie dużą pulę informacji rzeczywistych, którą następnie można wykorzystać do sprawdzenia użyteczności i poprawności skonstruowanego modelu (Przanowski, 2014).

W przypadku danych bankowych, sytuacja jest jeszcze trudniejsza z uwagi na wrażliwość informacji. Skutkuje to koniecznością występowania do instytucji finansowych z oficjalnymi podaniami, a otrzymane dane często są zafałszowane i zanonimizowane, co zwykle uniemożliwia ich zinterpretowanie. Znacząco utrudnia to tworzenie odpowiednio wiarygodnych modeli scoringowych, na co analitycy odpowiadają tworzeniem własnych, symulowanych danych (Przanowski, 2014).

Credit Scoring jest jednym z kluczowych narzędzi wykorzystywanych przez instytucje finansowe. W czasach gdy powszechny konsumpcjonizm i rozpędzone gospodarki świata generują potrzebę kreacji pieniądza, kredyt jest jedną z pierwszych rzeczy przychodzących na myśl. Zarówno wielkie korporacje, jak i osoby prywatne od czasu do czasu potrzebują znaczącego zastrzyku gotówki np. na kupno mieszkania, wymarzony samochód czy dobrze prosperującą inwestycję. Banki dziesiątki lat temu zauważyły potencjał leżący w tej dziedzinie analizy, a rozwój technologii może jeszcze bardziej zwiększyć niezawodność systemów predykcyjnych. Należy jednak pamiętać, że nawet najnowocześniejsze modele uczenia maszynowego potrzebują odpowiedniej puli wiarygodnych danych, aby najpierw skutecznie nauczyć się na błędach sprzed lat, a następnie zapobiec złym kredytobiorcom w przyszłości.

1.2 Uczenie Maszynowe

Karty punktowe, mimo że łatwe w interpretacji zarówno przez wnioskujących, jak i analityków tychże wniosków, nie są najbardziej optymalnym narzędziem do oceny wiarygodności kredytowej. Przez ostatnie kilkadziesiąt lat pojawiło się wiele nowych możliwości analizy, co jest związane z nieustającym rozwojem informatyzacji, a niektóre z nich zostały dopasowane do dziedziny Credit Scoring'u, dając lepszą efektywność predykcji oraz podnosząc zyski instytucji finansowych. W kolejnym podrozdziale skupiono się na opisie zagadnienia znanego jako Uczenie Maszynowe i jego wykorzystaniu w ocenie kredytowej.

1.2.1 Wprowadzenie do technik Uczenia Maszynowego

Uczenie Maszynowe, samouczenie się maszyn albo systemy uczące się, w języku angielskim tłumaczone jako Machine Learning (ML) jest dziedziną wchodzącą w skład nauk, zajmujących się Sztuczną Inteligencją (Artificial Inteligence – AI), Głównym jej celem jest tworzenie automatycznego systemu, który potrafi doskonalić się na bazie doświadczenia i nabywać na tej podstawie nową wiedzę. W uproszczeniu proces polega na znalezieniu wzorca w dostarczonych danych. Modele Uczenia Maszynowego powszechnie wykorzystywane są w wielu dziedzinach, których zachodzi potrzeba predykcji pewnego zjawiska (gov.pl, 2021).

Zadania ML ograniczone są do wąskiego, specyficznego zakresu, w którym ma działać dany system. W przeciwieństwie do sztucznej inteligencji, proces uczenia maszynowego nie jest w stanie stworzyć czegoś nowego, a jedynie uzyskiwać najbardziej optymalne rozwiązania w zadanym problemie. Najpopularniejszymi aplikacjami wykorzystującymi możliwości Uczenia Maszynowego są wyszukiwarki online, algorytmy podpowiadające najciekawsze dla użytkowników materiały w mediach społecznościowych, rozpoznawanie obrazów czy filtrowanie spamu ze skrzynek e-mail (elektronikab2b.pl, 2020).

Machine Learning stało się popularne relatywnie niedawno, ale jego historia jest znacznie dłuższa. Najważniejszy czynnik w analizie danych, czyli właśnie dane, zaczęto zbierać już w czasach starożytnych, a były to informacje o ilości zgromadzonej żywności, czy też szczegóły dot. spisów powszechnych. Kiedy z biegiem lat danych przybywało, ich analiza stawała się trudniejsza, a przełomowym momentem okazała się siedemnastowieczna epidemia dżumy, dziesiątkująca mieszkańców Anglii, gdy wówczas zaczęto publikować pierwsze zbiory danych dotyczące zdrowia publicznego. Jednakże przetwarzanie dużych zbiorów było procesem żmudnym, a jeszcze w drugiej połowie XIX wieku zebranie i przeanalizowanie

danych z przeprowadzonego w Stanach Zjednoczonych spisu powszechnego w zajmowało nawet dziesięć lat (fotc.com, 2022).

W latach 50. XX w. Alan Turing, znany ze swego wkładu w rozszyfrowanie Enigmy, niemieckiej maszyny szyfrującej, stwierdził, że „jeżeli maszyna będzie w stanie przekonać człowieka, że wcale nie jest maszyną, to będzie to świadczyć o osiągnięciu przez nią sztucznej inteligencji" (fotc.com, 2022). Z kolei Artur Samuel w latach 1952-1962 rozwijał program do szkolenia zawodników szachowych, jak również on, na konferencji w 1959 roku, po raz pierwszy użył pojęcia Uczenie Maszynowe jako „[…] dające komputerom możliwość „uczenia się" bez bycia konkretnie zaprogramowanym do danego zadania." (Mamczur, 2019). W roku 1957 Frank Rosenblatt opracował pierwszą komputerową sieć neuronową, która wczytując obrazy, generowała etykiety i kategoryzowała ilustracje (fotc.com, 2022). Kolejnym znanym systemem był Dendral z 1965 roku, który powstał na Uniwersytecie Stanforda z inicjatywy dwóch naukowców - Edwarda Feigenbauma oraz Joshuy Lederberga. Celem programu była automatyzacja analiz i identyfikacji molekuł związków organicznych nieznanych ówczesnym chemikom (britannica.com, 2019).

Przez kolejne dekady było dość cicho w zakresie Uczenia Maszynowego, a do lat dziewięćdziesiątych było wykorzystywane głównie w prostych grach. Następnie ML było coraz częściej wykorzystywane przede wszystkim przez przeglądarki internetowe tj. Google czy Yahoo oraz wspomagało systemy anty-spamowe. Wraz z początkiem drugiej dekady XXI wieku o Uczeniu Maszynowym ponownie zrobiło się głośno, głównie dzięki rozwojowi sieci neuronowych (Mamczur, 2019). Badanie prowadzone dziesiątki lat temu pozwoliły na to, by w trzeciej dekadzie XXI wieku, wiedza o Uczeniu Maszynowym nie była tak egzotyczna, a przeciętny student kierunku związanego z branżą informatyczną potrafił podać co najmniej kilka zastosowań ML w życiu codziennym. Jednakże aby modele oparte na algorytmach Machine Learning'u mogły być zastosowane w biznesie muszą być odpowiednio dostosowane do zagadnienia, a co za tym idzie – nauczone na bazie wprowadzonych danych. Uczenie maszynowe nie jest jednolitą technologią, a sposób jej działania zależy w dużej mierze od tego, z jakich algorytmów korzysta i jakimi danymi zostanie zasilona. Eksperci SAS wskazują 4 podstawowe techniki uczenia maszynowego (sas.com, 2018):

- Supervised Learning (ang. Uczenie Nadzorowane) - maszyny uczą się na podstawie dostarczonych przykładów, a dane wejściowe są wykorzystywane do wyszukiwania zależności, służących do rozwiązania określonego problemu. Gdy uda się ustalić

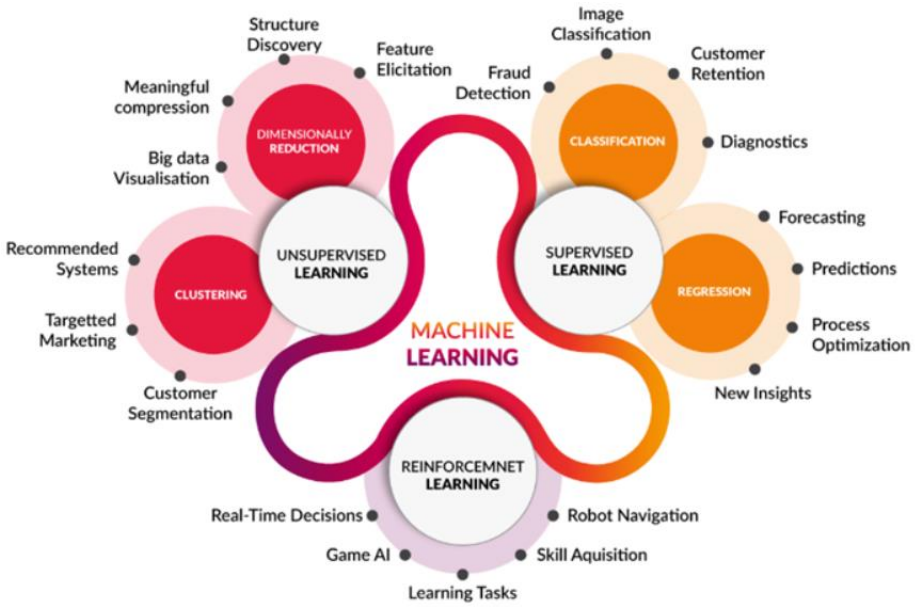
pewien wzorzec, jest on wykorzystywany w podobnych przypadkach. Do przykładowych zastosowań tej metody należą:

- zarządzanie ryzykiem;
- personalizacja interakcji;
- rozpoznawanie mowy, tekstu i obrazu;
- segmentacja klientów.

- Semi - Supervised Learning (ang. Uczenie Częściowo Nadzorowane) - maszyna otrzymuje zarówno dane wejściowe oznaczone (zawierające odpowiadające im dane wyjściowe, konkretne przykłady), jak i nieoznaczone (wymagające przyporządkowania do danych wyjściowych, znalezienia odpowiedzi). Ten rodzaj uczenia wykorzystuje się w sytuacjach, gdy dana instytucja dysponuje zbyt dużą ilością danych lub gdy informacje cechują się wysokim zróżnicowaniem, które uniemożliwia przyporządkowanie odpowiedzi do każdej z nich. W takiej sytuacji system sam proponuje odpowiedzi i jest w stanie stworzyć ogólne wzorce. Metoda znajduje zastosowanie w:
  - rozpoznawaniu mowy;
  - rozpoznawaniu obrazów;
  - klasyfikacji stron internetowych.
- Unsupervised Learning (ang. Uczenie Nienadzorowane) - maszyna nie posiada „klucza odpowiedzi" i musi sama analizować dane, szukać wzorców i odnajdować relacje. Ten typ ML najbardziej przypomina sposób działanie ludzkiego mózgu, który wyciąga wnioski na podstawie spontanicznej obserwacji i intuicji. Wraz ze zwiększaniem się rozmiaru zbiorów danych prezentowane wnioski są coraz bardziej precyzyjne. Poniżej przykłady wykorzystania:
  - analiza koszyka zakupowego;
  - wykrywanie anomalii;
  - rozpoznawanie podobnych obiektów.
- Reinforcement Learning (ang. Uczenie Wzmocnione) - maszyna otrzymuje gotowy zestaw dozwolonych działań, reguł i stwierdzeń oraz wykorzystuje reguły w taki sposób, aby osiągnąć pożądany efekt. Można to porównać do nauki gry np. w darta. Zasady określające, ile punktów musi zdobyć zawodnik oraz fakt zakończenia wartością podwójną pozostają niezmienne. Natomiast najoptymalniejsza kombinacja punktów otrzymanych z maksymalnie trzech rzuconych lotek zależy od indywidualnej decyzji gracza. Przykłady zastosowań to:

- nawigacja GPS (wybór trasy bazując na danych o natężeniu ruchu i pogodzie);
- przemysł gamingowy (dopasowanie scenariuszy rozgrywki do działań gracza);
- robotyka (dostosowanie natężenia pracy robotów do popytu).

Opisane techniki zilustrowano również na Rysunek 5:



Rysunek 5. Podstawowe techniki Uczenia Maszynowego *(Mamczur, 2019)*

W kolejnym podrozdziale szerzej opisano konkretne metody analizy danych wykorzystywane w dziedzinie Credit Scoring'u.

1.2.2 Metody Uczenia Maszynowego wykorzystywane w dziedzinie Credit Scoring

Dotychczas w bankowości nie stosowano powszechnie niektórych technik ML do zarządzania ryzykiem, a geneza takiego postępowania była zrozumiała – modele są trudne w interpretacji, a ponadto generują popyt na wysoce wyspecjalizowanych pracowników. Z drugiej zaś strony, rynek konkurencyjny zmienia się - transformacja cyfrowa, czy też nowy model bankowości otwartej wywierają wpływ na praktyki zarządzania ryzykiem. W tym kontekście stosowanie technik Uczenia Maszynowego zapewnia istotną przewagę, skracając czas podejmowania decyzji w procesach kredytowych oraz podnosząc ich skuteczność (crif.pl, 2018).

Główne metody, jakie stosuje się przy Credit Scoring'u należą do grupy technik klasyfikacji nadzorowanej (Bajek, 2011). W dalszej części pracy, celem dokładniejszego opisania, skupiono się na dwóch, najpopularniejszych metodach stosowanych do budowy modeli scoringowych, jakimi są regresja logistyczna i drzewa decyzyjne, pomijając równie ciekawe, aczkolwiek

rzadziej używane podejścia tj. sieci neuronowe, lasy losowe czy też SVM – Support Vector Machines (ang. Metoda Wektorów Nośnych) (StatSoft Polska, 2010).

Regresja logistyczna (LR) jest metodą statystyczną umożliwiającą ocenę wpływu wielu cech - tzw. zmiennych objaśniających - na szanse zajścia zdarzenia, np. zachorowania na pewną chorobę, czy też spłaty kredytu z planowanym terminie (lukaszderylo.pl, 2021). Model LR jest szczególnym przypadkiem uogólnionego modelu liniowego. Znajduje zastosowanie, gdy zmienna zależna jest dychotomiczna, to znaczy przyjmuje tylko dwie wartości takie jak na przykład sukces lub porażka, zwykle reprezentowane jako cyfry 1 i 0 (statystyka.az.pl, 2021).

Starając się jak najkrócej scharakteryzować to podejście, należy podkreślić, że głównym celem modeli regresji logistycznej jest znalezienie najlepszych współczynników(tzw. wag), które minimalizują błąd pomiędzy przewidywanym prawdopodobieństwem, a obserwowanym wynikiem. Realizowane jest to za pomocą algorytmu optymalizacji, takiego jak opadanie gradientu, celem dostosowania współczynników, aż model będzie odpowiednio pasował do danych, na których jest uczony, co następnie walidowane jest na zbiorze testowym (newsblog.pl, 2022). Wartości zmiennej dychotomicznej możemy przekształcić w postać prawdopodobieństwa wystąpienia danego zdarzenia, które przyjmuje wartości pomiędzy 0 lub 1. Gdy zastosuje się transformację logit możliwe jest zlinearyzowanie modelu regresji logistycznej i przedstawienie go w postaci regresji liniowej (naukowiec.org, 2014).

Regresja logistyczna jest najczęściej wykorzystywaną techniką modelowania scoringowego. Jej największą zaletą jest stabilność w czasie, co czyni ją względnie odporną na zaburzone dane. Może to dawać wrażenie, że metoda jest odporna na ataki, co zostało zweryfikowane na łamach tej pracy. Jej zaleta może być również interpretowana jako wada, ponieważ model może powodować nieoptymalność dopasowania do koniunktury rynkowej (Karolak, 2014).

Jako przykład implementacji modeli regresji liniowej można przedstawić badanie wpływu wysokości dochodów na fakt palenia papierosów przez badaną próbę osób. Przyjmijmy, że zmienna niezależna, jaką jest tu przypadku ilość dochodu, jest zmienną ilościową, a palenie papierosów określamy binarnie, gdzie 1 oznacza, że osoba pali papierosy, a 0 zdarzenie przeciwne. Jedna z takich analiz wskazała, że poziom dochodów jest istotny statystycznie, a osoby zarabiające więcej mają wyższą skłonność do palenia (naukowiec.org, 2014).

Drzewa decyzyjne (DT) są modelami o strukturze drzewiastej, podejmującymi decyzje na podstawie wartości poszczególnych cech. Można ich używać zarówno do klasyfikacji

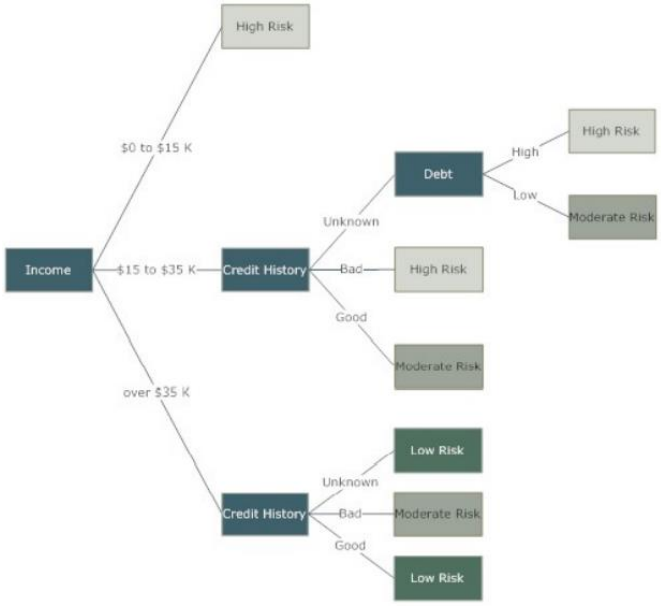
binarnej(służy do przewidywania, do której z dwóch klas\kategorii należy wystąpienie danych), jak i wieloklasowej (odnosi się do zadań klasyfikacyjnych, które mają więcej niż dwie etykiety klas). Pośród zalet drzew decyzyjnych należy wymienić prostotę w budowie i łatwość w interpretacji, dzięki czemu można dokładnie prześledzić cały proces podejmowania decyzji i jasno określić konkretne kryteria, na bazie których zrealizowano dany podział. Ich nauka jest szybka, a poza tym radzą sobie z zarówno liczbowymi jak i kategorycznymi typami danych (newsblog.pl, 2022). Są również bardzo skuteczne w przetwarzaniu znacznych ilości danych i nie wymagają ponadprzeciętnych mocy obliczeniowych (Bujak, 2008).

Nie wolno również zapomnieć o zagrożeniu jakie ze sobą niosą. Przy ich konstrukcji należy uważać na parametry, ponieważ jego nadmierna głębokość, czy też zbyt wiele utworzonych gałęzi może prowadzić do przeuczenia – nadmiernego dopasowania do danych treningowych – co spowoduje nieefektywne predykcje w środowisku produkcyjnym (newsblog.pl, 2022).

Drzewa decyzyjne mają ustalony porządek (Bujak, 2008):

- korzeń odpowiada wszystkim możliwym decyzjom;
- każdy wewnętrzny węzeł odpowiada pewnej decyzji, którą możemy podjąć;
- liściom odpowiadają cele.

Algorytm przedstawiono również na przykładzie prostej decyzji kredytowej na Rysunek 6:



Rysunek 6. Drzewo decyzyjne zastosowane do kategoryzacji potencjalnych kredytobiorców pod kątem ryzyka kredytowego *(Bujak, 2008)*

Rozwój Uczenia Maszynowego na początku XXI wieku, może dawać nadzieję na automatyzację w wielu branżach. Jesteśmy coraz bliżej powszechnego wykorzystania pojazdów autonomicznych, a od wielu lat modele ML są implementowane w dynamicznych start-up'ach, czy też w badaniach na uczelniach ekonomiczno-technicznych, stale udowadniając jak wszechstronne mogą być zastosowania systemów uczących się. Ze względu na ich mnogość i zróżnicowanie, poprzez odpowiedni dobór metod, analitycy podnoszą skuteczność stosowanych rozwiązań, dając wysoką satysfakcję z wysokiego poziomu poprawnych predykcji uzyskanych z modeli. Stajemy się coraz bardziej zależni od wyników otrzymywanych z analiz ML, czego przykładem są modele Credit Scoring. Dziś to od maszyny zależy, czy wnioskodawca otrzyma kredyt i zrealizuje marzenie o mieszkaniu czy nowym telewizorze. Jednakże wraz ze wzrostem roli maszyn w naszym codziennym życiu, pojawia się pytanie – czy są one bezpieczne? Temat odporności systemów Uczenia Maszynowego został poruszony w kolejnym podrozdziale.

1.3 Adversarial Machine Learning – Wrogie Uczenie Maszynowe

Modele Uczenia Maszynowego otworzyły zupełnie nowe możliwości w dziedzinie automatyzacji, a wizja wszechobecnej Sztucznej Inteligencji skutecznie rozpala wyobrażenia ludzi o świecie, w którym będziemy na porządku dziennym wykorzystywać roboty, posiadające własną świadomość. Jednakże należy pamiętać, że z wielką mocą wiąże się wielka odpowiedzialność, a wykorzystanie nowych możliwości w złym celu, może nieść ze sobą groźne skutki. W kolejnym podrozdziale pochylono się nad problemem Wrogiego Uczenia Maszynowego, czyli potencjalnych ataków na modele ML.

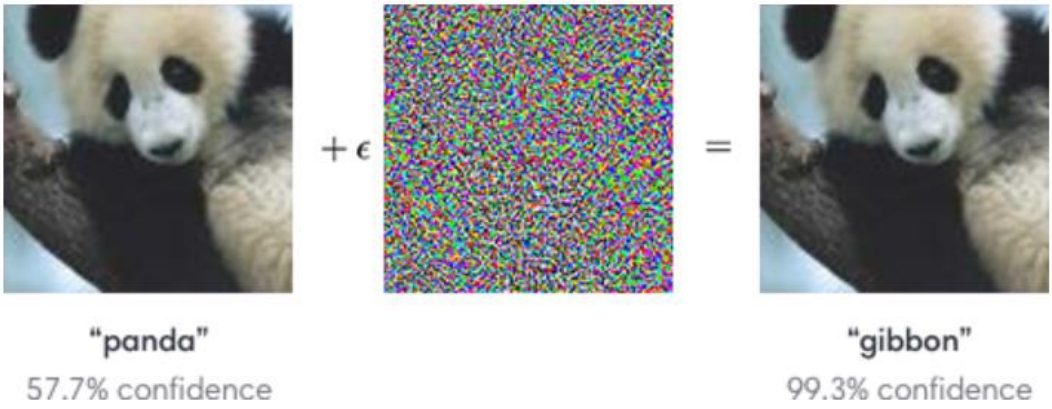
1.3.1 Charakterystyka Wrogiego Uczenia Maszynowego

Adversarial Machine Learning (AML), tłumaczone na język polski jako Kontradyktoryjne Uczenie Maszynowe, czy też Wrogie Uczenie Maszynowe to dziedzina badań, która koncentruje się na opracowywaniu modeli ML odpornych na ataki kontradyktoryjne, które są stosowane do oszukiwania lub manipulowania modelami uczenia maszynowego przez wprowadzanie złośliwych lub wprowadzających w błąd danych podczas faz uczenia lub wnioskowania. Ataki te mogą mieć poważne konsekwencje, od kradzieży poufnych informacji po nieprawidłowe działanie krytycznych systemów, takich jak samojezdne samochody lub urządzenia medyczne.

W celu dokładniejszego opisania zjawiska AML, w kolejnych akapitach odwoływano się nie do ataków na konkretne systemy Uczenia Maszynowego, lecz w ujęciu ogólnym – jako ataki na Sztuczną Inteligencję (AI). W tej tematyce wkraczamy również w dziedzinę bezpieczeństwa w cyberprzestrzeni. W sektorze cyberbezpieczeństwa AML próbuje oszukać modele poprzez tworzenie unikalnych, wprowadzających w błąd danych wejściowych, aby zmylić model, powodując jego wadliwe działanie (zephyrnet.com, 2022).

Przeciwnicy (w języku angielskim nazywani są jako „attackers”, tłumaczone na język polski również jako „napastnicy”) mogą wprowadzać dane, które mają za zadanie zmanipulować rezultaty wyjściowe, wykorzystując luki w modelu. Nie jesteśmy w stanie zidentyfikować tych danych wejściowych ludzkim okiem, jednak powoduje to nieprawidłowe działanie modelu. W systemach AI występują różne formy podatności, takie jak tekst, pliki audio, obrazy. Dużo łatwiej przeprowadzać ataki cyfrowe, takie jak manipulowanie tylko jednym pikselem w danych wejściowych, co może prowadzić do błędnej klasyfikacji (zephyrnet.com, 2022).

Przykład manipulacji obrazu przedstawiono na Rysunek 7. W tym przypadku zaatakowano system rozpoznawania zwierząt, który został nauczony na bazie pewnej puli zdjęć. Przed atakiem, model rozpoznał, że na fotografii znajduje się panda, określając to z pewnością bliską 58%. Po dodaniu szumu, zmanipulowano system do tego stopnia, iż z niemal 100% przekonaniem sklasyfikował zdjęcie pandy jako gibbona (openai.com, 2017). Łatwo zauważyć, że jednym zaburzeniem napastnik zmienił status modelu z przydatnego na bezużyteczny.



Rysunek 7. Przykład ataku na system rozpoznawania obrazów *(openai.com, 2017)*

O ile zmanipulowanie systemu rozpoznawania obrazów zwierząt może wydawać się niegroźne i niedostatecznie ukazywać niebezpieczeństwo płynące z tego typu ataków, o tyle wpływ napastników np. na działanie samochodów autonomicznych wskazuje na tragiczne skutki jakie mogą zostać spowodowane. Jednym z mniej skomplikowanych algorytmów ML zastosowanych w tych środkach transportu jest system rozpoznawania znaków drogowych, jako że ich liczba jest skończona i względnie nieduża, a ich kształt, kolor i rozmiar jest ściśle znormalizowany. Dla przykładu rozważmy typową sytuację drogową.

Rysunek 8. Widok z kamery przedniej samochodu autonomicznego. Właściwie rozpoznany znak STOP *(Jerzy Surma, 2020)*

Na Rysunek 8 zamieszczono widok z przedniej kamery samochodu autonomicznego, tuż przed skrzyżowaniem. Auto bez problemu rozpoznaje znak STOP, a następnie wykonuje odpowiednie czynności, aby zatrzymać się przed skrzyżowaniem. Jednakże bardzo łatwo jest wprowadzić system w błąd, co może wydarzyć się na skutek zabrudzenia znaku, czy pomalowania go farbą, czego przykład przedstawiono na Rysunek 9 (Jerzy Surma, 2020).

Rysunek 9. Widok z kamery przedniej samochodu autonomicznego. Niewłaściwie rozpoznany znak STOP *(Jerzy Surma, 2020)*

Wskutek tego typu zaburzenia danych wejściowych, system oparty na Uczeniu Maszynowym może nie tylko nie rozpoznać tego znaku jako nakaz zatrzymania się, a wręcz może przypisać do otrzymanego obrazu zupełnie inny znak, mający w danym momencie krytyczne znaczenie dla bezpieczeństwa kierowcy. Na Rysunek 10 przedstawiono przykładową interpretację przez model ML, gdzie zabrudzony znak STOP został przyjęty jako znak pierwszeństwa (Jerzy Surma, 2020).

Rysunek 10. Widok z kamery przedniej samochodu autonomicznego. Znak STOP błędnie rozpoznany jako znak bezwzględnego pierwszeństwa przy skręcie w lewo *(Jerzy Surma, 2020)*

Zakładając, że aby dotrzeć do celu kierowca na danym skrzyżowaniu musi skręcić w lewo, samochód bez zatrzymywania się przejedzie przez to skrzyżowania. Skutki takiej decyzji z wysokim prawdopodobieństwem będą tragiczne w skutkach, co ukazuje jak duże znaczenie ma jakość danych dostarczanych do modelu i jak niewielkie zaburzenie może wpłynąć na jego niezawodność, a co za tym idzie, uzasadnienie dalszego wykorzystania w biznesie (Jerzy Surma, 2020).

1.3.2 Typy i przykłady ataków na algorytmy Uczenia Maszynowego

Na daną chwilę większość algorytmów proponowanych przez badaczy, naukowców i specjalistów z branży R&D skupia się głównie na wysokiej wydajności i niskiej liczbie błędnych klasyfikacji. Jednakże nawet kiedy te cele zostają osiągnięte, modele te często nie powinny być implementowane w środowiskach produkcyjnych, zwłaszcza w domenach krytycznych, czy aspektach życia, które mogą mieć wpływ na życie znacznej części społeczeństwa, nie uwzględniając innych kryteriów i wymagań dotyczących sztucznej inteligencji. Są nimi: bezpieczeństwo algorytmów, ich interpretowalność i uczciwość. Co więcej, rezultaty osiągane są na danych, które są odpowiednio przygotowane w warunkach laboratoryjnych i możliwe jedynie gdy implementacja też zachodzi w takich warunkach (Pawlicki, 2020).

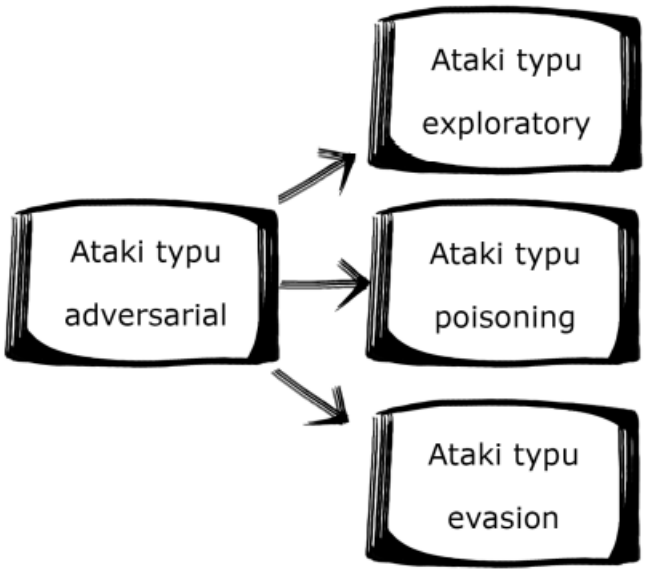
Zastosowanie Sztucznej Inteligencji na wielką skalę stało się rzeczywistością, za czym idzie świadomość, że bezpieczeństwo samych algorytmów Uczenia Maszynowego wymaga natychmiastowej uwagi. Adwersarze, jak również nazywa się atakujących systemy ML, potrafią starannie dobrać próbki danych wejściowych, aby zmieniało to wyniki klasyfikacji lub regresji w oczekiwany przez nich sposób. Świadomość zagrożeń związanych z ich użyciem, a także ich podatność na AML jest jeszcze całkiem niewielka (Pawlicki, 2020), jednakże już powstały definicje określające poszczególne typy ataków.

Jednym ze znanych podziałów jest klasyfikacja ze względu na dostęp do modelu (zephyrnet.com, 2022):

- Atak białoskrzynkowy - odnosi się do sytuacji, w której atakujący ma pełny dostęp do modelu docelowego. Obejmuje to architekturę i parametry, które pozwalają im tworzyć próbki danych na modelu docelowym. Osoby atakujące będą miały ten dostęp tylko wtedy, gdy testują model jako programista. Mają oni detaliczną wiedzę na temat architektury sieci oraz znają tajniki modelu i tworzą strategię ataku;

- Atak czarnoskrzynkowy - odnosi się do sytuacji, w której atakujący nie ma dostępu do modelu docelowego i może jedynie zbadać dane wyjściowe.

Podział ataków na czarnoskrzynkowe i białoskrzynkowe oparty jest o umiejscowienie atakującego. Inna klasyfikacja bazuje na strategii ingerencji w model, a jej schemat przedstawiono na Rysunek 11.



Rysunek 11. Nowe ataki na Uczenie Maszynowe *(Pawlicki, 2020)*

Atak zatruwający (typu poisoning) koncentruje się na danych ze zbioru uczącego. Tutaj atakujący zmienia istniejące lub wprowadza nieprawidłowo oznakowane dane. Wskutek takiego działania, model przeszkolony na „zatrutym" zbiorze będzie tworzył błędne predykcje na prawidłowo oznakowanych danych (Towards Data Science, 2021). W literaturze znaleźć można kilka artykułów na temat ataków tego typu. W jednym z nich, autorzy opisują użycie tzw. wrogiego szumu etykiet (ang. adversarial label noise). W tym artykule przedstawiona jest metoda wykorzystująca sposób działania algorytmu SVM (Support Vector Machines – ang. Metoda Wektorów Nośnych), którego działanie polega na mapowaniu danych na wielowymiarową przestrzeń właściwości w sposób umożliwiający kategoryzację punktów danych (IBM, 2021).

Ogólnym założeniem ataku jest wprowadzenie do zbioru treningowego próbki, która znacząco zmieni wynik klasyfikacji, obniżając skuteczność modelu. Taką próbkę można stworzyć poprzez rozwiązanie problemu optymalizacyjnego, polegającego na wyszukiwaniu lokalnych maksymów powierzchni funkcji błędu, do czego wykorzystano algorytm gradient ascent. Atak

wykorzystuje odwracanie etykiet konkretnych próbek w klasyfikacji binarnej zbioru uczącego, przy założeniu, że dane w zbiorze walidacyjnym nie są w żaden sposób zmieniane (Battista Biggio, 2012).

Ataki unikowe (typu evasion), w odróżnieniu od ataków zatruwających, nie skupiają się na danych używanych do uczenia modelu, lecz na odpowiedniej manipulacji danymi wejściowymi, dla których model wydaje prognozowany rezultat. Polegają one na modyfikowaniu danych, aby wydawały się uzasadnione, lecz by prowadziły do błędnej prognozy. Przykładem wykorzystania tego typu ataków są modele oceny wiarygodności kredytowej. Ubiegając się o kredyt, osoba atakująca może zamaskować swój prawdziwy kraj pochodzenia za pomocą usługi VPN, ukrywając w ten sposób np. iż jest obywatelem kraju uznawanego przez model jako bardziej ryzykowny, co mogłoby zmniejszyć jego szanse na pozytywną ocenę jego wniosku (Towards Data Science, 2021).

Innym kierunkiem wykorzystania ataków unikowych są modele służące do odfiltrowywania wiadomości e-mail będących spamem. Ich podejście może polegać na eksperymentowaniu z mailami, które model już wytrenował w zakresie sprawdzania i rozpoznawania jako spam. Jeśli model został wyszkolony do filtrowania wiadomości e-mail zawierających konkretne słowa, atakujący może tworzyć nowe e-maile zawierające powiązane z tym słowa, które przejdą przez algorytm, co spowoduje, że wiadomość e-mail, która w typowym procesie zostałaby sklasyfikowana jako spam, spamem nie jest, co w oczywisty sposób pogarsza skuteczność modelu (zephyrnet.com, 2022).

Atak poszukiwawczy (typu exploratory) może wystąpić po wytrenowaniu algorytmu, a jego zadaniem jest odkrywanie informacje o wewnętrznym działaniu modelu, w celu identyfikacji słabych punktów. W tym podejściu ingerencja jest ukierunkowana na poszukiwanie informacji o (Yi Shi, 2017):

- granicy decyzyjnej używanej przez algorytm (np. hiperpłaszczyzny maszyny wektorów nośnych (SVM) algorytm);
- ogólnym zestawie reguł, którymi kieruje się algorytm;
- zestawie logicznych lub probabilistycznych właściwości algorytmu;
- danych, które zostały wykorzystane (lub nie wykorzystane) do uczenia algorytmu.

W ostatnich latach powstało kilka prac, badających ataki na Głęboką Sieć Neuronową (DNN). W artykule z 2017 roku (Yi Shi, 2017) naukowcy zbudowali tzw. funkcjonalny ekwiwalent

klasyfikatorów modelu DNN, opierając się na algorytmach SVM oraz naiwnego klasyfikatora Bayes'a. Z kolei w badaniu z 2019 roku (Shi Y., 2019) opisano jak przy pomocy Głębokiej Sieci Neuronowej można wydobyć klasyfikator wdrożony warunkach produkcyjnych. Podsumowując - celem ataku typu exploratory jest budowa lokalnego klasyfikatora (Pawlicki, 2020).

Jak zatem zabezpieczyć się przed wrogimi ingerencjami w model? Jedną z możliwości stanowi tzw. trening kontradyktoryjny, będący jednym z podstawowych podejść do poprawy wydajności i bezpieczeństwa ML. Polega on na kontrolowanym atakowaniu modelu na wiele sposobów, znajdując w ten sposób „czułe punkty" zastosowanego rozwiązania. Na skutek takich badań, osoby budujące model mogą zweryfikować jego odporność na wrogie ataki, a następnie podjąć odpowiednie kroki celem poprawy jego bezpieczeństwa (zephyrnet.com, 2022).

Adversarial Machine Learning jest względnie nową poddziedziną Analizy Danych i na tę chwilę nie łatwo jest znaleźć rzeczywiste przykłady ingerencji w rzeczywiste, używane produkcyjnie modele ML. Dzięki wysokiej świadomości analityków nie trzeba uczyć się na nieświadomie popełnionych błędach następnie wykorzystanych przez adwersarzy, a problem został zidentyfikowany zanim pojawiły się jego potencjalnie fatalne skutki. W związku z powyższym, w ostatnich latach, naukowcy prowadzą intensywne badania, samodzielnie generując różne scenariusze ingerencji w model, jednakże wciąż pozostaje w tej dziedzinie wiele przestrzeni dla nowych odkryć.

Podsumowanie rozdziału

[PODSUMOWANIE ROZDZIAŁU]

2 Wprowadzenie do budowy modelu drzewa decyzyjnego w Credit Scoring

[WSTĘP DO ROZDZIAŁU]

2.1 Osiągnięcia w dziedzinie wykorzystania drzew decyzyjnych w Credit Scoring

[WSTĘP DO PODROZDZIAŁU]

2.1.1 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

2.1.2 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

2.1.3 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

2.2 Wybór i opis wykorzystanego zbioru danych

[WSTĘP DO PODROZDZIAŁU]

2.2.1 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

2.2.2 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

2.2.3 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

2.3 Koncepcja przeprowadzenia części praktycznej

[WSTĘP DO PODROZDZIAŁU]

2.3.1 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

2.3.2 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

2.3.3 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

Podsumowanie rozdziału

[PODSUMOWANIE ROZDZIAŁU]

3 Budowa modelu drzewa decyzyjnego do celu Credit Scoring

[WSTĘP DO ROZDZIAŁU]

3.1 Wykorzystane narzędzia i technologie

[WSTĘP DO PODROZDZIAŁU]

3.1.1 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

3.1.2 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

3.1.3 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

3.2 Budowa modelu

[WSTĘP DO PODROZDZIAŁU]

3.2.1 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

3.2.2 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

3.2.3 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

3.3 Analiza wyników

[WSTĘP DO PODROZDZIAŁU]

3.3.1 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

3.3.2 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

3.3.3 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

Podsumowanie rozdziału

[PODSUMOWANIE ROZDZIAŁU]

4 Atak na opracowany model

[WSTĘP DO ROZDZIAŁU]

4.1 Strategia badania odporności modelu

[WSTĘP DO PODROZDZIAŁU]

4.1.1 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

4.1.2 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

4.1.3 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

4.2 Implementacja wybranych technik ataku

[WSTĘP DO PODROZDZIAŁU]

4.2.1 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

4.2.2 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

4.2.3 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

4.3 Analiza wyników i weryfikacja hipotez badawczych

[WSTĘP DO PODROZDZIAŁU]

4.3.1 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

4.3.2 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

4.3.3 Tytuł podpodrozdziału

[TEKST PODPODROZDZIAŁU]

Podsumowanie rozdziału

[PODSUMOWANIE ROZDZIAŁU]

Wnioski

Bibliografia

1. Bajek, R. (2011). *WYKORZYSTANIE METOD EKSPLORACJI DANYCH DO BUDOWY MODELI SCORINGOWYCH.* Politechnika Śląska, Instytut Informatyki.

2. bankier.pl. (2012, kwiecień 2). *Tajemnicza liczba, czyli credit scoring.* Pobrano z lokalizacji Bankier.pl: https://www.bankier.pl/wiadomosc/Tajemnicza-liczba-czyli-credit-scoring-2512458.html

3. Battista Biggio, B. N. (2012). *Poisoning Attacks against Support Vector Machines.*

4. britannica.com. (2019, wrzesień 19). *DENDRAL.* Pobrano z lokalizacji Britannica: https://www.britannica.com/technology/DENDRAL

5. Bujak, Ł. (2008). *Drzewa decyzyjne.* Pobrano z lokalizacji is.umk.pl: http://www.is.umk.pl/~duch/Wyklady/CIS/Prace%20zalicz/08-Bujak.pdf

6. crif.pl. (2018). *Rola „machine learning" w procesach kredytowych.* Pobrano z lokalizacji CRIF: https://www.crif.pl/wiadomo%C5%9Bci/dla-prasy/2020/sierpie%C5%84/rola-machine-learning-w-procesach-kredytowych/

7. Dean Caire, S. B. (2006). *A handbook for developing credit scoring systems in a microfinance context.* Washington: Development Alternatives, Inc.

8. direct.money.pl. (2022, styczeń 18). *direct.money.pl.* Pobrano z lokalizacji Co to jest scoring kredytowy? Jak banki ustalają credit scoring i jakich używają systemów?: https://direct.money.pl/artykuly/porady/czym-jest-credit-scoring

9. elektronikab2b.pl. (2020, grudzień 11). *Czym jest uczenie maszynowe i jak można je wykorzystać?* Pobrano z lokalizacji elektronikab2b.pl: https://elektronikab2b.pl/biznes/53039-czym-jest-uczenie-maszynowe-i-jak-mozna-je-wykorzystac

10. Encyklopedia Zarządzania. (2020, maj 19). *Credit Scoring.* Pobrano z lokalizacji Encyklopedia Zarządzania: https://mfiles.pl/pl/index.php/Credit_scoring

11. experian.com. (2021, luty 11). *What Is a Good Credit Score?* Pobrano z lokalizacji Experian: https://www.experian.com/blogs/ask-experian/credit-education/score-basics/what-is-a-good-credit-score/

12. forbes.com. (2021, czerwiec 28). *What Is A Good Credit Score?* Pobrano z lokalizacji Forbes: https://www.forbes.com/advisor/credit-score/what-is-a-good-credit-score/

13. fotc.com. (2022, listopad 16). *Machine Learning — czym jest uczenie maszynowe?* Pobrano z lokalizacji FOTC: https://fotc.com/pl/blog/machine-learning/

14. gov.pl. (2021, czerwiec 15). *Co to jest uczenie maszynowe – inteligentna analiza danych?* Pobrano z lokalizacji gov.pl: https://www.gov.pl/web/popcwsparcie/co-to-jest-uczenie-maszynowe--inteligentna-analiza-danych

15. IBM. (2021, sierpień 17). *Sposób działania algorytmu SVM.* Pobrano z lokalizacji ibm.com: https://www.ibm.com/docs/pl/spss-modeler/saas?topic=models-how-svm-works

16. Jerzy Surma, d. h. (2020). *Prezentacja pt. Hakowanie Sztucznej Inteligencji.* Warszawa: Szkoła Główna Handlowa.

17. Karolak, Z. (2014, listopad 18). *Dynamiczne ujęcie ryzyka kredytowego z ujęciem analizy przeżycia.* Pobrano z lokalizacji sas.com: https://www.sas.com/content/dam/SAS/pl_pl/image/events/mdb/prezentacje/zuzanna-karolak-dynamiczne-ujecie-ryzyka.pdf

18. lukaszderylo.pl. (2021, styczeń 31). *Regresja logistyczna - co to jest?* Pobrano z lokalizacji Łukasz Deryło: https://www.lukaszderylo.pl/blog/regresja-logistyczna.html

19. Mamczur, M. (2019, listopad 30). *Czym jest uczenie maszynowe? I jakie są rodzaje?* Pobrano z lokalizacji Mirosław Mamczur: https://miroslawmamczur.pl/czym-jest-uczenie-maszynowe-i-jakie-sa-rodzaje/

20. naukowiec.org. (2014, kwiecień 14). *Regresja logistyczna - opis.* Pobrano z lokalizacji Naukowiec.org: https://www.naukowiec.org/wiedza/statystyka/regresja-logistyczna_466.html

21. newsblog.pl. (2022, grudzień 18). *Wyjaśnienie regresji a klasyfikacja w uczeniu maszynowym.* Pobrano z lokalizacji News Blog: https://newsblog.pl/wyjasnienie-regresji-a-klasyfikacja-w-uczeniu-maszynowym/#Regresja_logistyczna

22. openai.com. (2017, luty 24). *Attacking Machine Learning with Adversarial Examples.* Pobrano z lokalizacji OpenAI: https://openai.com/blog/adversarial-example-research/

23. Pawlicki, M. (2020). *Zastosowanie Metod Uczenia Maszynowego do Wykrywania Ataków Sieciowych.* Bydgoszcz: Uniwersytet Technologiczno-Przyrodniczy im. Jana i Jędrzeja Śniadeckich.

24. pl.economy-pedia.com. (2021). *Punktacja kredytowa.* Pobrano z lokalizacji pl.economy-pedia: https://pl.economy-pedia.com/11030209-credit-scoring

25. Przanowski, K. (2014). *Credit Scoring w erze Big-Data.* Warszawa: Szkoła Główna Handlowa.

26. researchgate.net. (2017, październik). Pobrano z lokalizacji ResearchGate: https://www.researchgate.net/figure/The-5V-of-Big-Data-Characteristics_fig1_321050765

27. sas.com. (2018, sierpień 22). *Cztery typy uczenia maszynowego.* Pobrano z lokalizacji SAS: https://www.sas.com/pl_pl/news/informacje-prasowe-pl/2018/cztery-typy-uczenia-maszynowego.html

28. scoringexpert.pl. (2018, luty 21). *8 ważnych informacji potrzebnych do zrozumienia nowej oceny punktowej, którą BIK sprzedaje konsumentom.* Pobrano z lokalizacji Scoring Expert: http://scoringexpert.pl/2018/02/21/ocena-punktowa-bik-skala-do-100/

29. Shi Y., S. Y. (2019). *Generative Adversarial Networks for Black-Box API Attacks with Limited Training Data. .*

30. Siddiqi, N. (2016). *Credit Risk Scorecards Developing and Implementing Intelligent Credit Scoring.* New Jersey: John Wiley & Sons, Inc.

31. StatSoft. (2010). *Zastosowanie metod scoringowych w działalności bankowej.* Pobrano z lokalizacji StatSoft: https://media.statsoft.pl/_old_dnn/downloads/zast_met_skoringowych_w_dz_bankowej.pdf

32. StatSoft Polska. (2010). *Metody skoringowe w biznesie i nauce.* Pobrano z lokalizacji media.statsoft.pl: https://media.statsoft.pl/_old_dnn/downloads/modele_skoringowe_w_biznesie.pdf

33. statystyka.az.pl. (2021, sierpień 2017). *Regresja Logistyczna.* Pobrano z lokalizacji Statystyka od A do Z: https://www.statystyka.az.pl/regresja-logistyczna.php

34. techtarget.com. (2021, marzec). *5 V's of big data.* Pobrano z lokalizacji techtarget.com: https://www.techtarget.com/searchdatamanagement/definition/5-Vs-of-big-data

35. The World Bank Group. (2019). *Credit scoring approaches guidelines.* Washington: The World Bank Group.

36. Thomas L.C., E. D. (2002). *Credit Scoring and Its Applications.* Philadelfia: Society for Industrial and Applied Mathematics.

37. Thonabauer G., N. B. (2004). *Guidelines on Credit Risk Management. Credit Approval Process and Credit Risk Management.* Oesterreichische Nationalbank and Austrian Financial Market Authority.

38. totalmoney.pl. (2020, październik 3). *Ocena punktowa BIK-u – jaka wartość scoringu BIK-u jest dobra i daje szansę na kredyt?* Pobrano z lokalizacji Totalmoney: https://www.totalmoney.pl/artykuly/179147,kredyty-gotowkowe,ocena-punktowa-bik-u---jaka-wartosc-scoringu-bik-u-jest-dobra-i-daje-szanse-na-kredyt,1,1

39. Towards Data Science. (2021, lipiec 12). *What is Adversarial Machine Learning?* Pobrano z lokalizacji towardsdatascience.com: https://towardsdatascience.com/what-is-adversarial-machine-learning-dbe7110433d6

40. Weston, L. (2012). *Your Credit Score.* New Jersey: Pearson Education, Inc.

41. Wikipedia. (2022, grudzień 24). *FICO.* Pobrano z lokalizacji Wikipedia: https://en.wikipedia.org/wiki/FICO.

42. Wysiński, P. (2013). *Zastosowanie scoringu kredytowego w bankowości.* Gdańsk: Uniwersytet Gdański.

43. Yi Shi, Y. S. (2017). *How to Steal a Machine Learning Classifier with Deep Learning.* Rockville, MD 20855, USA: Intelligent Automation, Inc.,.

44. zephyrnet.com. (2022, marzec 3). *Co to jest kontradyktoryjne uczenie maszynowe?* Pobrano z lokalizacji Zephyrnet: https://zephyrnet.com/pl/co-to-jest-kontradyktoryjne-uczenie-maszynowe/

Spis rysunków

Spis tabel

Załączniki

Streszczenie w języku polskim

[WPISAĆ STRESZCZENIE PRACY]

Oświadczenie autora pracy

[WKLEIĆ GOTOWĄ FORMATKĘ]