Nowoczesne Systemy Zarządzania
Zeszyt 19 (2024), nr 2 (kwiecień-czerwiec)
ISSN 1896-9380, s. 87-112
DOI: 10.37055/nsz/200432

Instytut Organizacji i Zarządzania
Wydział Bezpieczeństwa, Logistyki i Zarządzania
Wojskowa Akademia Techniczna
w Warszawie

Modern Management Systems
Volume 19 (2024), No. 2 (April-June)
ISSN 1896-9380, pp. 87-112
DOI: 10.37055/nsz/200432

Institute of Organization and Management
Faculty of Security, Logistics and Management
Military University of Technology
in Warsaw

Jak wykorzystać modele Markowa na potrzeby przeciwdziałania praniu pieniędzy oraz w walce z terroryzmem? (Część 1)

How to use Markov models for the purposes of counteracting money laundering and in the fight against terrorism? (Part 1)

Maciej Aleksander Kędzierski
Radca prawny, Okręgowa Izba Radców Prawnych, Warszawa, Polska
sulawezi.mk@onet.eu; ORCID: 0000-0003-3074-1355

Abstrakt. Współczesne zjawiska przestępcze, takie jak między innymi terroryzm i jego finansowanie, jedynie w określonym zakresie bazują pod względem wykonawstwa na schematach. W rzeczywistości sprawcy działają niekonwencjonalnie, zmieniają taktykę postępowania, a także schematy samych komórek terrorystycznych. Tym samym coraz trudniej jest prowadzić rozpoznanie tej przestępczej aktywności, gdyż wiedza o niej jest ograniczona (ukryta), a ponadto następuje czasowe opóźnianie w zakresie analizy przyczynowo-skutkowej. Tym samym sięga się po nowe rozwiązania z zakresu matematyki. Jednym z nich jest stosowanie ukrytych modeli Markowa, których właściwości umożliwiają budowanie modeli predykcyjnych opartych na ograniczonych zasobach wiedzy, cechowaniu elementów oraz ocenie prawdopodobieństwa zdarzeń. Ich zastosowanie jest szerokie i dotyczy nie tylko rozpoznawania fizycznych aktów terroryzmu, lecz także ich finansowania czy procederu prania pieniędzy.
Słowa klucze: ukryte modele Markowa, pranie pieniędzy, finansowanie terroryzmu, stan ukryty (ukryty model), proces rozpoznania anomalii

Abstract. Contemporary criminal phenomena, such as terrorism and its financing, are based on patterns only to a certain extent in terms of execution. In reality, perpetrators act unconventionally, changing the tactics of action but also the patterns of the terrorist cells themselves. Thus, it is increasingly difficult to conduct reconnaissance of this criminal activity, because knowledge about it is limited (hidden) and, in addition, there is a time delay in the cause-effect analysis. Thus, new solutions from the field of mathematics are reached for. One of them is the use of hidden Markov models, the properties of which allow building predictive models based on limited knowledge resources, characterization of elements and assessment of the probability of events. Their application is wide and concerns not only the recognition of physical acts of terrorism, but also their financing or money laundering.
Keywords: Hidden Markov Models (HMM), money laundering, financing terrorism, hidden state (hidden model), anomaly recognition process

Wprowadzenie

Zmieniająca się rzeczywistość w obszarze usług finansowych, rozwój sztucznej inteligencji czy zwiększanie systemowe obowiązków instytucji obowiązanych (dalej jako: IO) w zakresie przeciwdziałania praniu pieniędzy oraz finansowaniu terroryzmu (dalej jako: AML/CFT) wymusza poszukiwanie coraz nowszych instrumentów wsparcia procesów analitycznych i decyzyjnych w tych instytucjach. Poszukiwanie nowych rozwiązań analitycznych związane jest z dużą różnorodnością samego zjawiska finansowania terroryzmu oraz podobieństw do zachowań sprawców prania pieniędzy. Zwłaszcza widoczne jest to w kontekście zmiany negatywnych zachowań dwubiegunowych na rzecz rozproszonego środowiska inicjatorów i wykonawców, co nie sprzyja kontroli sposobów i taktyki przestępczego postępowania. Aktualny stan zagrożenia wynika z triady zagrożeń, która składa się z: upadłych państw (np. Liban/Somalia/Jemen), rozprzestrzeniania broni masowego rażenia oraz globalizacji terroryzmu. Sam zaś terroryzm (TR) charakteryzuje się: znaczną atrakcyjnością wykorzystywania metod terrorystycznych jako sposobów uzyskiwania celów, poszerzeniem się kręgu wsparcia koncentrującego się na akolitach i prywatnych sponsorach terroryzmu oraz znaczącą dywersyfikacją metod i zachowań zwiększających aktywa finansowe i militarne dla beneficjentów terrorystycznych. Nadal metody terrorystyczne i zbliżone do nich działania sabotażowe zastępują lub wręcz prowokują, niedopuszczając do bezpośrednich konfliktów zbrojnych o charakterze wojennym realizowanych w nieświadomości dla potencjalnych podmiotów finansujących tego typu pośrednie siłowe rozwiązania. Ponadto upadłe państwa mogą generować negatywne skutki uboczne transnarodowego terroryzmu daleko poza własne granice.

Na potrzeby kryminalistyczne systemu przeciwdziałania ML/FT budowane są rozwiązania mające na celu zidentyfikowanie ukrytych powiązań składających się z państw wspierających terroryzm, aktywnej komunikacji wewnątrz grup i organizacji terrorystycznych, komunikacji pomiędzy członkami grup terrorystycznych a podmiotami ze wsparcia otoczenia czy też między nimi a reprezentantami przestępczości zorganizowanej. Ale także określanie ram czasowych aktu planowania aktów terrorystycznych oraz zmapowanych stanów zagrożenia, pozwalających na podjęcie właściwej kontrreakcji.

Kreowanie czynności analitycznych, w ramach systemów zarządzania informacją w IO, opartych na doświadczeniu i zdolnościach pracowników pionu AML/CFT, poszerzone zostało już w latach 90. XX w. o podstawowe programy analityczne oceny sieci, jak np. iBase (i2 Analyst's Notebook). Tu także należałoby zwrócić uwagę na zindywidualizowane narzędzie analityczne opracowywane na potrzeby konstruowania modeli zdarzeń predykcyjnych, tak aby instytucje mogły nie tylko przewidywać zagrożenie (ryzyko), lecz i być przygotowane na skuteczną ich neutralizację (obniżanie kosztów ryzyka).

W obszarze tym pomocne stało się wykorzystanie modeli Markowa, których możliwości spowodowały powstanie ukierunkowanych badań także na rzecz doskonalenia przeciwdziałania zagrożeniom w obszarach ryzyka AML/CFT. Modele Markowa wykorzystywane są do opracowywania prawdopodobieństw różnych stanów, ale także w zakresie oceny szybkości zmian między poszczególnymi wzorcami, w zależności od przyjętych warunków. Niejednokrotnie obserwowalne stają się takie zachowania sprawców ML/FT, polegające na tym, że sprawca kreuje pewien indywidualny model przygotowania lub realizacji zachowania przestępczego (lub też korzysta z takiego modelu jako już wykreowanego), a także aby dokonać przestępstwa – musi skorzystać z określonego modelu narzuconego jako usługa/produkt przez daną IO. W ten sposób powstają zindywidualizowane modele przestępczego postępowania lub też schematy postępowania, które uwidaczniają się jako anomalie w systemach wewnętrznych IO. Tym samym niezbędne staje się ustalenie określonej sekwencji danych, nawet gdy dane te początkowo będą niepełne lub nie wszystkie te dane będą dostępne dla IO, którymi w dalszej kolejności będzie można posłużyć się dla konstrukcji wzorca.

Dane te ogólnie będzie można podzielić na te, które są uzyskiwane w wyniku aktywności podmiotów systemu AML/CFT (w tym samych IO) oraz na inne dane pozasystemowe (np. plasowane w mediach społecznościowych, uzyskiwane w ramach kierunkowej filtracji Internetu czy jako informacje z systemu geolokalizacji). Wydaje się, że w takich sytuacjach możliwe jest skorzystanie między innymi z modeli Markowa. Co istotne, modele takie można wprowadzić także na potrzeby analizowania aktywności terrorystycznej, np. w mediach społecznościowych. Tym samym określone metody statystyczne będzie można wykorzystać szerzej niż tylko na potrzeby analizy danych, które generują zwłaszcza relacje klient/transakcja – IO (wąskie źródło informacji), także z wykorzystaniem modeli szeregów czasowych opartych na ciągu obserwacji pewnego zjawiska w kolejnych jednostkach czasu.

Szeregi czasowe stanowią podstawę do analizy dynamiki zjawisk powiązanej z prognozowaniem zachowań. Cecha dynamiczna pozostaje wskazana również wobec potrzeby prowadzenia dynamicznej oceny ryzyka IO i to w relacji z klientem/transakcją (Słowniczek, pkt 3). Zastanawiając się nad możliwością zastosowania modeli Markowa na potrzeby AML/CFT, należałoby mieć na uwadze to, że stanowią one narzędzie modelowania statystycznego wykorzystywane do analizy i przewidywania zjawisk o charakterze sekwencji zdarzeń. Podejście oparte na modelach Markowa jest niejako uzupełnieniem dotychczasowych rozwiązań w obszarze AML/CFT, które budowane były na podstawie reguł, skali ryzyka (scoring), klastrów prawdopodobieństw czy wzorców postępowania. Ponadto należy mieć na uwadze to, iż mimo wielu źródeł informacji zbudowanie z nich całościowej sieci powiązań przestępczych wydaje się bardzo trudne i obarczone „lukami informacyjnymi", stąd też próby ich wdrożenia mogą dotyczyć zwłaszcza „zachowań lokalnych", czyli charakterystycznych dla danego analizowanego obszaru.

Dotyczy to zarówno budowania sieci powiązań opartych na analizie kryminalnej, jak i powiązań transakcyjnych wobec rozpoznawania zagrożeń w ramach systemu AML/CFT. W konsekwencji należy poszukiwać takich instrumentów, które pozwolą analitykom na uzupełnienie tej wiedzy. Możliwości takich dostarczają między innymi instrumenty probabilistyczne (oparte na prawdopodobieństwie). Modele Markowa to uzyskiwanie możliwości pomiędzy projektowaniem schematów taktycznych walki z FT i poszukiwaniem w obszarze zagrożeń takich czynników cechujących negatywne zachowania, które następnie będą mogły posłużyć do budowy modeli (wzorców) kontrdziałań wykrywczych i kreowania operacji specjalnych, w których podobnie jak w schematach sieciowych, będą wyłapywane aktywności przestępcze. Stąd też potrzeba sprzężenia schematów przestępczych z modelami matematycznymi na potrzeby uzyskiwania efektów skutecznych przeciwdziałań. Przy czym o ile nie będzie możliwe wykrycie poszczególnego skonkretyzowanego przypadku FT, to ważnym efektem będzie już uzyskanie wniosku na rzecz ujawnienia możliwego schematu predykcyjnego działania przestępczego. Subsumpcja takiego schematu umożliwiać będzie na dalszym etapie prowadzonego rozpoznania operacyjnego ustalanie sprawców czy też podejmowanie czynności identyfikacyjnych „podejrzalności" w IO w ramach analizy rozpoznania.

Przedmiotowy artykuł (przygotowany został w dwóch częściach) stanowi próbę oceny możliwości wykorzystania metody ukrytych modeli Markowa do rozpoznawania negatywnych zjawisk wiążących się ze współczesnym terroryzmem oraz sposobami jego finansowania, a także przeciwdziałania praniu pieniędzy. Przedmiotem artykułu nie jest ocena porównania metod Markowa wobec innych modeli analitycznych stosowanych w obszarze AML/CFT. W tym przypadku istotne jest wskazanie na samą metodę, a zwłaszcza na jedną z odmian, tj. zastosowanie ukrytych modeli Markowa. Dlatego też w części pierwszej opracowania wskazano na wykorzystanie HMM (ang. Hidden Markov Model) wobec oceny ujawnienia sieci sprawców ML/FT na podstawie zachowań transakcyjnych oraz na podstawie wzajemnych relacji pomiędzy grupami/sprawcami. W części drugiej zaś zaprezentowano wykorzystanie HMM jako instrumentu do ujawniania ukrytych (grup) sprawców przestępstw prania pieniędzy do identyfikacji ryzyka/zagrożenia oraz ujawniania potencjalnych terrorystów poza systemem AML/CFT. Jako przesłankę do podjęcia badań należy wskazać na możliwość wykorzystania HMM, jako metody wczesnego wykrycia zagrożenia związanego z aktywnością sprawcy ML/TR, a także próby budowania przeciwdziałań na podstawie uzyskanego rezultatu, jakim jest schemat predykcyjnego/możliwego zachowania sprawcy. Jednocześnie do rozważenia pozostaje wykorzystanie HMM na rzecz oceny zachowań klientów/sprawców IO, które będzie można ocenić jako spełniające warunek wstępny do oceny „podejrzalności" lub jako procesu „podejrzalnego", który będzie można zakwalifikować przez cel pośredni, czyli np. okoliczności z art. 74 lub art. 86 ustawy z dnia 1 marca 2018 r. o przeciwdziałaniu praniu pieniędzy oraz finansowaniu terroryzmu (dalej jako: ustawy o p.p.p.f.t.) (Ustawa, 2018)

w ramach systemu AML/CFT. W tym celu – jako metody badawcze – dokonano przeglądu dotychczasowych badań, które prezentowały wyniki i metodykę postępowania z wykorzystaniem HMM na potrzeby ujawnienia informacji o podejrzanych transakcjach. Dlatego też w tak określonym zakresie artykuł stanowi materiał przeglądowy. Założeniem jest, że przedmiotowe rozwiązanie korzystające z na tyle uniwersalnej metody, jaką jest HMM, pozwala na jej zastosowanie także w IO, wobec potrzeby wyłonienia z wielości cechowanych danych – metadanych – jako wzorców postępowania będących określonymi jako negatywny impuls (anomalia kierunkowa).

Należy zaznaczyć, że przepisy ustawy o p.p.p.f.t. są jedynie przepisami ogólnymi, kierunkowymi i na tym poziomie kształtującymi obowiązki IO. W konsekwencji, na tak określonym obszarze funkcjonowania IO, powinny realizować obowiązki AML/CFT w sposób przez siebie ustalony i w granicach wskazanego prawa. Posłużenie się określoną metodą do realizacji celu – ustalenia wielkości ryzyka i sformułowania adekwatnych wobec niego środków bezpieczeństwa finansowego (Słowniczek, pkt 9) – w znacznej mierze jest autonomiczną decyzją IO. Metody te zaś jednocześnie powinny być efektywne i kompatybilne z całością systemu AML/CFT, łącznie z zadaniami wykonywanymi przez poszczególną IO. W takim zakresie HMM zwłaszcza mógłby wesprzeć proces rozpoznania anomalii kierunkowej, gdy nie ma możliwości rzeczywistego wyszukiwania takich wzorców, a możliwe jest jedynie ich „losowe" pozyskiwanie z zastosowaniem rozkładu prawdopodobieństw i określenia macierzy prawdopodobieństw przejść między stanami. Wobec zastosowania metody opartej na HMM możliwe byłoby także jej zastosowanie, gdy w danej IO oraz pomiędzy nimi (także różnymi rodzajowo) mamy do czynienia jedynie z „porwanym", czyli niekompletnym łańcuchem dostaw przestępczych aktywów (tylko niektóre stany są znane). Jako metodę badawczą wykorzystaną w publikacji zastosowano przegląd literatury i badań na temat możliwości HMM oraz próby wdrożenia jej na potrzeby AML/CFT.

W konsekwencji celem publikacji jest przybliżenie i zaprezentowanie HMM jako pewnego rozwiązania pozwalającego na ocenę ryzyka ML/FT wobec ograniczonych danych/informacji o źródle i aktywności wykonawczej sprawcy zagrożenia. O możliwości zastosowania HMM przesądza możliwość wykreowania wiedzy o dalszym kroku sprawcy na podstawie ograniczonej liczby stanów oraz na wykorzystaniu podejścia probabilistycznego opartego na rozbudowanych i kierunkowych zbiorach uczących. HMM jako metoda badawcza wobec AML/CFT jest sprawdzalna, mimo wielu mankamentów, w sytuacjach, w których występują szerokie braki wiedzy o zdarzeniu terrorystycznym, jego planowaniu, a budowanie przeciwdziałań na podstawie danych historycznych nie sprawdza się w zakresie ustalania działań predykcyjnych.

Samo podejście tematyczne należałoby zaliczyć do multidyscyplinarnego, które można byłoby przyrównać do kryminalistyki statystycznej (obliczeniowej). Schemat ocenny wychodzi znacznie poza zasięg tradycyjnej kryminalistyki, które zdecydowanie bada dane i ślady pozostawione (w wyniku zaistniałych zdarzeń, faktów).

W omawianym przypadku możliwe jest także typowanie rodzaju i miejsca przypuszczalnego, przyszłego pozostawienia śladów przestępczej działalności (predykcja śladów). Należy zauważyć, że w kontekście przedstawianego tematu występuje także pojęcie *terrorism informatics* – informatyka terrorystyczna (Chen, Reid, Sinai et al., 2008; Słowniczek, pkt 4), które koncentruje się głównie na wykorzystaniu analizy sieci społecznościowych (SNA, ang. Social Network Analysis) do strukturalnej i pozycyjnej analizy sieci terrorystycznych, w przypadku których wymagane informacje są dostarczane z danych niezwiązanych z przestępczością. SNA bada relacje – powiązania i zależności – pomiędzy elementami wewnątrz grup społecznych. Poza tym prezentowane podejście może posłużyć do profilowania działań zaradczych w przypadku zarządzania bezpieczeństwem w ramach systemu AML/CFT, a także wobec prowadzenia czynności operacyjno-rozpoznawczych jednostek współpracujących (np. policji, służb specjalnych). HMM powinno umożliwić przewidywalność kolejnych zachowań w ramach taktyki przestępczej sprawców ML/FT działających dotychczas w danych ustalanych stanach.

Wykorzystanie ukrytych modeli Markowa na rzecz walki z przestępczością terrorystyczną – ogólny zarys

Mając na uwadze to, że incydenty terrorystyczne występują rzadko, co utrudnia ich ocenę z punktu widzenia maszynowego uczenia się i wnioskowania za pomocą modeli (wzorców), przedmiotem poszukiwania powinny być inne rozwiązania. Ponadto w zakresie identyfikacyjnym występuje także problem niejednoznaczności czasowej, brakujących danych i niejednoznaczności atrybucyjnej, stąd też potrzeba poszukiwania danych w tych obszarach, które mogą aktywizować się częściej niż na potrzeby dokonywania aktów terrorystycznych, np. w czasowych zakresach przygotowań logistycznych, werbunkowych, zaspokajania potrzeb (Raghavan, Galstyan, Tartakovsky, 2013, s. 2402-2430). Dlatego też powstały propozycje rozwiązań, aby w zakres budowania profili zachowań wsparcia terrorystycznego wprząc między innymi ukryte modele Markowa (HMM, ang. Hidden Markov Models) i to w obszarze finansowania działań terrorystycznych oraz budowania otoczki logistycznej i informacyjnej tych przestępczych przedsięwzięć. Należy mieć na uwadze czynniki typujące, takie jak:

– finansowanie terroryzmu związane jest nie tylko z finansowaniem aktu terrorystycznego, lecz także innych przejawów aktywności terrorystów (dodatkowo dla udowodnienia przestępstwa – finansowania terroryzmu – nie jest niezbędne faktyczne dostarczenie środków dla decydenta, ale ważna jest subsumpcja zachowania sprawcy wobec strony przedmiotowej przestępstwa);

- sfera finansowa związana jest także np. z podejmowaniem działań również po dokonaniu zamachu (jako integralnie z nim związanych, dla przykładu przekazanie środków rodzinie terrorysty samobójcy), a także w czasie pomiędzy poszczególnymi atakami terrorystycznymi (aktami terroru), np. na działalność logistyczną, propagandową czy szkoleniową – tym samym powstaje znacznie więcej symptomów/śladów (cech zachowań), z których będzie można budować zbiór uczący, uwzględniając również szeregi czasowe (podejście takie zakłada, że ocena finansowania zachowań okołoterrorystycznych dostarcza więcej „śladów podejrzalności" niż jedynie śledzenie skonkretyzowanego procesu dostarczania aktywów na rzecz samego zamachu terrorystycznego);
- na potrzeby ustalania danych możliwe jest wykorzystanie nie tylko informacji o przebiegu procesu dostarczania środków, lecz także źródeł ich pochodzenia (kryminalne, pozakryminalne, państwowe) oraz danych wygenerowanych z otwartych źródeł informacji;
- łańcuchy dostaw aktywów na rzecz terrorystów przebiegają nie tylko w ramach świadczenia usług przez podmioty nienadzorowane (pozasystemowe), lecz także przez różnorodne rodzajowo podmioty-uczestników systemu AML/CFT, jako mających status IO – stąd możliwość pozostawiania wielości danych, w celu ich dalszego przetwarzania, także przez analizę szeregów czasowych dystrybucji aktywów, gdy determinowane są przez strukturę zapisów i identyfikacji ich finansowej dystrybucji niezależnie od tego, czy plasowane są one u beneficjenta terrorystycznego, czy nie;
- uzupełnieniem dla danych pozyskiwanych w ramach systemu AML/CFT mogą być dane operacyjne służb otrzymywane ze źródeł działań pozasystemowych (operacyjnych), dlatego też poszerza to podejście do zakresu danych celem przetworzenia (dane te mogą być niejednorodne, ale uzupełniające), zwłaszcza może mieć to miejsce przy wykorzystaniu „czynnika ludzkiego" do budowania zbiorów uczących;
- mimo dążności do nasycenia decydenta/analityka informacjami z różnych źródeł niekoniecznie uzyskiwany jest pełny obraz, a wręcz luki informacyjne są zjawiskiem towarzyszącym zbiorom informacji o funkcjonowaniu ugrupowania i stanie zaangażowania przygotowań do samego zamachu (należy w tym zakresie wyzbywać się szumu informacyjnego i dokonywać walidacji danych przed ich zastosowaniem w modelu).

Jednocześnie samo środowisko informacyjne – wobec TR – charakteryzuje się tym, że model przeciwdziałania musi być w stanie (1) wykrywać potencjalne zagrożenia terrorystyczne w bardzo „zaśmieconym" środowisku, (2) prowadzić efektywnie analizę dużej ilości danych oraz (3) generować hipotezy z użyciem jedynie informacji częściowych i niedoskonałych (Pattipati, Willett, Allanach et al., 2006, s. 27-50).

Ta cecha ukrycia w modelach Markowa staje się priorytetowa z punktu widzenia ujawnienia warstwy aktywności przestępczej, ale także posłużenia się nią na potrzeby niejawnych działań operacyjnych służących do neutralizacji struktur terrorystycznych (jak np. podjęcie działań wobec uzyskanych ustaleń probabilistycznych).

Jedną z podstawowych argumentacji, jaką należałoby uwzględnić przy potrzebie wręcz sięgnięcia w zakresie AML/CFT do HMM, jest ta, która łączy się również z zasadą wykorzystania metod nauk matematycznych w innych środowiskach niż matematyczne, np. w naukach społecznych czy w naukach o bezpieczeństwie. Tak więc odpowiednio opracowany model matematyczny pozwala przewidzieć przebieg zjawiska lub zachowanie się układu fizycznego w różnych warunkach, bez konieczności przeprowadzania eksperymentów praktycznych (Fabian, Szedel, 2002, s. 302). To podejście wiele wyjaśnia, dlaczego stosowana jest HMM jako teoretyczna metoda badawcza wobec ML/FT (TR) i jak wiele ona umożliwia w procesie rozpoznawania anomalii, do czego zobowiązane są IO w ramach wykonywania swoich ustawowych obowiązków. Niejednokrotnie monitoring klienta, aby stwierdzić jego potencjalną możliwość podejrzenia udziału w procederze ML/FT (PZ/TR), wykonywany jest poza jego wiedzą, deskryptywnie. Istotne jest to, aby IO już wyprzedzająco zapewniła sobie w czasie t pozyskiwania niezbędnych danych dla modelowania czynników o kliencie/transakcji, tak aby IO mogła budować jak najwłaściwsze parametry (z myślą o użyciu środków bezpieczeństwa finansowego) na potrzeby HMM (np. dla ogólnych czynników: uzyskiwanie wiedzy ze szczegółowych prognoz informacji, tj. demograficznych, przestrzennych i czasowych, wydarzeń związanych z niepokojami społecznymi).

Należy też zwrócić uwagę na to, że aby wynik budowany na metodzie HMM zapewnił właściwy efekt – istotny jest odpowiedni dobór topologii (liczby stanów i rodzaju powiązań między stanami), elementów dla zbiorów uczących oraz wartości parametrów modelu (Figielska, 2011, s. 64). Dlatego też z punktu widzenia dalszego zastosowania HMM w zakresie rozpoznania ML/FT niezbędna staje się ekstrakcja i kwantyfikacja danych przez IO w uświadomieniu czemu (jakiemu procesowi zarządzania nimi i decydowania na podstawie uzyskanych wyników) te przetworzone kierunkowo dane będą służyły. Zwłaszcza aby stwierdzić, jaki sposób zarządzania informacją pozostaje jak najbardziej adekwatny wobec rodzaju uzyskanego wyniku. Wynik może być wykorzystany zarówno dla oceny ryzyka instytucjonalnego IO, jak i na potrzeby oceny indywidualnego ryzyka generowanego przez klienta/transakcję. Do rozwiązania tego problemu należałoby rozważyć użycie testu na danych symulacyjnych. Test mógłby odpowiedzieć na to, w jakim zakresie czasowym może wystąpić anomalia i określić sytuację, która w jakiś sposób odbiegałaby od pozostałych. Niemniej ważne będzie zaprojektowanie struktury danych wartości i wypełnienie jej danymi wyznaczonymi opartymi na celu postępowania (cel – struktura monitorowania klienta/transakcji i stosowania środków bezpieczeństwa finansowego).

Dane typowane do wykorzystania mogą być kolejno estymowane i trenowane za pomocą algorytmów Bauma–Welcha (np. na potrzeby optymalizacji parametrów, Słowniczek, pkt 1) i Viterbiego (informacja w dalszej części artykułu, Słowniczek, pkt 2). W tym przypadku istotne jest uzyskanie największej wiarygodności, tj. wyznaczenia takich wartości parametrów, które maksymalizowałyby prawdopodobieństwo wygenerowania przez model określonych wartości zmiennej objaśnianej, jaką można byłoby zaobserwować w rzeczywistości. Algorytm Bauma–Welcha wykorzystuje się w procesie uczenia HMM. Sam proces uczenia polega na takim zmodyfikowaniu macierzy, by zmaksymalizować prawdopodobieństwo wygenerowania zbioru sekwencji uczących. Struktura modelu (liczba stanów, przejścia między stanami i wyprowadzane w nich symbole) jest w tym przypadku z góry narzucona (Przybysz, Kasprzak, 2012).

Mając na uwadze przedstawioną argumentację, można wskazać, że dopiero teraz – gdy poszerzono system przeciwdziałania praniu pieniędzy o działania wymierzone w finansowanie terroryzmu, przy jednoczesnym zadaniowaniu wielości podmiotów w poszczególnych krajach, w których funkcjonuje system AML/CFT – wprowadziło to podstawy do możliwości pozyskiwania wielości danych do ich przetwarzania na potrzeby badań statystycznych. Dotychczasowe badania opierające się jedynie na danych związanych z faktem dokonanych zamachów były niewystarczające i obejmowały długi czas. Długość okresów badawczych nie wydaje się w tym przypadku stanem sprzyjającym, ponieważ organizacje terrorystyczne często zmieniają (także intertemporalnie pomiędzy zamachami) zarówno taktykę działania, jak i obszary geograficzne swojej aktywności, które są specyficzne dla danego regionu. Tak więc możliwe pomyłki mogą wystąpić w przypadku uśredniania danych, co wymusza badanie każdego obszaru osobno (z uwzględnieniem jego specyfiki). Tym samym już z założenia może nastąpić niejednorodność oceny danych. Przyjmując natomiast założenie dla badania jedynie aktywności danego wytypowanego podmiotu terrorystycznego, niekoniecznie można uzyskać szerszy globalny obraz aktywności terrorystów, a tym samym przyjęte wnioski przy przenoszeniu ich na inne pozaprawnie działające podmioty może powodować wystąpienie błędu co do oceny (chociażby przez bazowanie na danych historycznych cechujących jedynie daną pojedynczą organizację terrorystyczną). Tego typu cechę, zindywidualizowaną, można wykorzystać na potrzeby identyfikacji sprawców aktów terrorystycznych, w przypadku gdy do nich żadna organizacja się nie przyznaje lub też przyznaje się, ale inna organizacja niż ta, która rzeczywiście dokonała tych aktów. Organizacja może działać podobnie, tj. w ogólności jak inne podmioty, a w szczegółach na przykład taktycznych różnić się od pozostałych. Podobne błędy mogą wystąpić także przy wykorzystaniu jedynie otwartych źródeł informacji (OSINT, ang. *open-source intelligence*), jako zewnętrznego zasilania informacyjnego dla analityka (tu także trzeba brać pod uwagę dane przekazywane w wyniku działań dezinformacyjnych).

Dlatego też, o ile będzie to możliwe, przed wprowadzeniem danych z OSINT do zbioru uczącego powinny one zostać poddane walidacji (a przynajmniej gradacji ich wiarygodności). Wielość codziennych aktywności terrorystycznych i wzajemnych powiązań pozostaje ukryta, a nawet nieustalona, tym samym niezbędne staje się wprowadzenie takich badań statystycznych, które mogłyby zaradzić lukom informacyjnym czy też proponować scenariusze predykcyjne.

Inną trudnością może być to, że stosując OSINT, można także bazować na danych nieprawdziwych, w sposób świadomy wprowadzonych do przestrzeni medialnej (świadoma dezinformacja). Dlatego też wyzwaniem staje się zastosowanie, w tym przypadku, między innymi ukrytych modeli Markowa. Nadal jednak problemem pozostają zmienne z parametrami wybranymi w celu optymalizacji metryki (jako parametry przyjętego wzoru), stanowiące podstawy do dalszych czynności obliczeniowych. Model może stać się „modelem weryfikacyjnym" i ustaleniowym zdarzeń (jako najbardziej prawdopodobnych) wobec możliwości bazowania na danych fałszywych. Niemniej po oczyszczeniu przedpola danych bez ich interpretacji umożliwią one zastosowanie HMM, co pozostaje sednem sprawy. To ten model Markowa ma mieć decyzyjne znaczenie wobec i tak niezależnie od niego występującego fałszu informacyjnego, dezinformacji czy szumu informacyjnego. Ten stan wiedzy jest tożsamy także z innym rodzajem działań analitycznych. Podejście probabilistyczne, krótki łańcuch typowania stanów i właściwe cechowanie danych jest głównym atutem tego modelu.

Tak więc coraz bardziej skomplikowane pod względem wielości i struktury dane generowane przez IO w systemie AML/CFT powodują determinację w zakresie poszukiwania nowych rozwiązań z dziedziny matematyki i sztucznej inteligencji na potrzeby wspierania procesów analitycznych i decyzyjnych w IO. Jedną z propozycji wstępnego otwarcia na tworzenie podstaw do takich rozwiązań jest wspomniany proces Markowa. Andriej Markow to rosyjski matematyk przełomu XIX i XX w., który zajmował się w swoich obszarach badawczych zwłaszcza teorią liczb, analizą matematyczną czy rachunkiem prawdopodobieństwa. Jego rozwiązania określane jako łańcuchy Markowa, proces Markowa czy własność Markowa dały asumpt do dalszego rozwoju matematyki zajmującej się rachunkiem prawdopodobieństwa. Mimo że w czasach Markowa nie znano pojęcia „prania pieniędzy", to współcześnie wielokrotnie naukowcy, przygotowując różne modele rozwiązań w obszarze analizy i decyzyjności wobec potrzeby ujawnienia nieprawidłowości związanych z tym przestępczym procederem, wykorzystują te pojęcia, których pierwotnym autorem w matematyce był Markow. Proces Markowa jest procesem stochastycznym (Słowniczek, pkt 8), czyli przeciwieństwem procesu deterministycznego i opiera się na prawdopodobieństwie zdarzeń. Stąd też jako punkt wyjścia nie jest potrzebny dla uzyskania danych wtórnych wobec jakiegoś faktycznie zaistniałego zdarzenia (aktu terrorystycznego) badanie dalekiej przeszłości zachowań sprawców, ale podstawą do dalszych analiz matematycznych może być także grupowanie danych na podstawie jakiejś niezaobserwowanej dynamiki organizacji, do czego można się odnieść przez ukryty model Markowa (Fabian, Szedel, 2002, s. 302-303).

W reprezentującym stanie układu nie są znane wartości zmiennych losowych. Znane są jedynie wartości zmiennych losowych innego procesu stochastycznego Markowa. Proces stochastyczny $\{X_t\}_t$ określany jest procesem Markowa, jeżeli dla każdego momentu t_0 prawdopodobieństwo dowolnego położenia systemu w przyszłości ($t > t_0$) zależy tylko od jego położenia w chwili $t = t_0$ i nie zależy od tego, w jaki sposób proces ten przebiegał w przeszłości. Mówimy, że jest to proces „bez pamięci" (Filipowicz, 1996, s. 27). Jest to więc model, w którym następuje takie zgrupowanie potencjalnych okazji, że prawdopodobieństwo każdej okazji zależy tylko od stanu osiągniętego w poprzednim wydarzeniu. Ukryty model Markowa to skończony zbiór stanów, z których każdy jest powiązany z rozkładem prawdopodobieństwa. Przy tak rozumianym procesie istotne jest założenie, że stan bieżący zależy tylko od skończonej liczby stanów poprzednich. Czyli sam wspomniany proces to taki, w którym stan bieżący zależy tylko od stanu poprzedniego i nie zależy od innych stanów wcześniejszych. Dla przykładu ocena działalności terrorystycznej organizacji może być wystawiona na podstawie pewnych stanów/cech grupy, zamiast całej przeszłej historii funkcjonowania tej grupy (niejednokrotnie także nieznanej). W odróżnieniu od łańcuchów Markowa, w przypadku których zmiany stanów mogą następować tylko w całkowitych momentach czasu $t = 1, 2,...$, w procesie Markowa zmiany stanów mogą następować w dowolnych momentach czasu $0 < t_1 < t_2 < ... < +\infty$. Dla dowolnych $0 < t < u < +\infty$, $i, j = 1, 2,... , p$ tym samym definiujemy prawdopodobieństwa przejścia ze stanu s_i w momencie t do stanu s_j w momencie u jako prawdopodobieństwa warunkowe $\pi_{ij}(t, u) := P(X_u = s_j \mid X_t = s_i)$. Proces Markowa nazywa się jednorodnym łańcuchem Markowa, jeżeli $\pi_{ij}(t, u)$ zależy tylko od i, j oraz od różnicy $u - t$ (zob. Łochowski, 2019). Tak więc w dyskretnych momentach ukryte procesy znajdują się w pewnym stanie i generowana jest ich obserwacja. Następnie ukryty proces ulega zmianie, jego stan następuje na bazie prawdopodobieństwa przejścia. Głównym celem jest więc ujawnienie (ukrytych) stanów z zastosowaniem właściwości ukrytego łańcucha Markowa, przy określonej sekwencji obserwacji. W konsekwencji takie właściwości pozwalają zarządzać informacją przetworzoną (jako mówiącą o czymś) wobec stosunkowo niewielkiej liczby danych, co w konsekwencji przyspieszyć powinno proces decyzyjny w IO.

Istnieją cztery rodzaje modeli Markowa, które mogą być wykorzystane sytuacyjnie:

- **łańcuch Markowa** – modeluje stan systemu za pomocą zmiennej losowej, która ewoluuje w czasie. Jej rozkład zależy wyłącznie od rozkładu poprzedniego stanu;
- **ukryty model Markowa, HMM** – obserwacje są związane ze stanem systemu, ale są one zazwyczaj niewystarczające do jego dokładnego określenia, gdyż jest on tylko częściowo obserwowalny. Ukryte modele Markowa różnią się od klasycznych łańcuchów Markowa brakiem możliwości bezpośredniej obserwacji stanu, w jakim przebiega proces. Zamiast tego obserwujemy

realizację probabilistycznej funkcji określonej na zbiorze stanów procesu (Mazurek, 2010, s. 27);
- **procesy decyzyjne Markowa** – przejścia między stanami zależą od stanu bieżącego i wektora akcji, który jest stosowany do danego systemu; używany do obliczania strategii działań, która maksymalizuje pewną przydatność w odniesieniu do oczekiwanych korzyści;
- **częściowo obserwowalne procesy decyzyjne Markowa** – stan systemu jest tylko częściowo obserwowalny, jednak system jest w pełni kontrolowany. Dlatego też ostatnie metody przybliżania uczyniły go użytecznym w wielu zastosowaniach, takich jak sterowanie prostymi robotami (BusinessInsider, 2021).

Można więc rozpatrywać przyszłe, przypuszczalne zdarzenia jako wynik oceny stanu na podstawie stanu poprzedniego lub też stanu na podstawie pewnego ustalonego procesu (funkcji probabilistycznej). Opisując potrzeby w zakresie AML/CFT, należy zwrócić uwagę na ukryte modele Markowa. Są to modele probabilistyczne, w których stan procesu opisuje pojedyncza dyskretna zmienna losowa X_t. HMM stanowi pewne rozszerzenie definicji łańcucha Markowa. Algorytmy HMM mogą być stosowane w wielu dziedzinach, w jakich celem jest odnalezienie ciągu danych, których stan nie jest obserwowany, ale inne zależne od tej sekwencji dane (stany) są obserwowane. Stąd też wobec zastosowania HMM w AML/CFT niezbędne jest takie nakierowanie uzyskiwania danych, że z jednej strony będzie można uzyskać wiedzę o danych „ukrytych" na potrzeby ustalenia procederu przestępczego, a z drugiej badanie „widocznych" danych, które o tych ukrytych danych będą mogły coś powiedzieć. Przy ich zastosowaniu system jest przedstawiany jako proces Markowa o niewidocznych dla obserwatora stanach, ale z widocznym wyjściem (obserwacją), które jest losową funkcją stanu. HMM ma skończony zbiór stanów, z których każdy jest powiązany z (ogólnie wielowymiarowy) rozkładem prawdopodobieństwa. Przejścia między stanami są zdefiniowane przez zbiór prawdopodobieństw zwanych prawdopodobieństwami przejścia. Co oznacza prawdopodobieństwo przyjmowania przez kolejne stany sygnału zadanych wartości przy założeniu, że właściwości statystyczne sygnału nie zmieniają się w czasie. W konkretnym stanie generowany jest wynik lub obserwacja, zgodnie z powiązanym prawdopodobieństwem dystrybucji. Z zewnątrz widoczny jest tylko wynik, a nie stan obserwatora i dlatego stany są ukryte na zewnątrz, stąd też nazwa – ukryty model Markowa (Fabian, Szedel, 2002, s. 304).

Na potrzeby prowadzenia obliczeń niezbędne staje się uzyskanie ze zbioru danych informacji, które umożliwią określenie: długości ciągu obserwacji, zbiór stanów, stan początkowy, macierz przejścia między stanami, macierz prawdopodobieństwa emisji obserwacji czy liczbę symboli w ciągu obserwacji. Obserwacje w każdym czasie są niezależne pod warunkiem znajomości ukrytego stanu X_t. Oznacza to, że obserwacja w czasie t zależy tylko od ukrytego stanu w czasie t, co bardzo często odnosi się do założenia o lokalnej niezależności, która jest głównym założeniem całej grupy modeli ze zmiennymi ukrytymi (Genge, 2014). Należy jednak wskazać,

że zasadniczym problemem występującym przy budowie ukrytych modeli Markowa, jak już przedstawiono, jest odpowiedni dobór topologii (liczby cech/stanów i rodzaju powiązań między stanami) oraz wartości parametrów modelu, który zapewniałby jego dobre działanie, np. wysoką zdolność rozpoznawania (Fabian, Szedel, 2002, s. 304; Figielska, 2011, s. 64). Wytypowanie rodzajów stanów ML/FT w IO należy już do analityka, który podejmując tę decyzję, powinien także mieć na uwadze możność zidentyfikowania cech i rodzajów powiązań. Należy jednak pamiętać o tym, że w większości przypadków HMM to uczenie maszynowe bez nadzoru powoduje określone konsekwencje. Do nich należą takie następstwa, iż trudne jest poszukiwanie i identyfikowanie rodzaju wzorców (np. na podstawie przebiegów czasowych o zmiennym i losowym charakterze) oraz nie ma jakiejś oczywistej miary błędu. Stąd też w przypadku podjęcia prac badawczych działania mogą mieć wyłącznie na celu zbadanie możliwości zastosowania HMM jako metody do wykorzystania jej do wyciągania wniosków w określonych obszarach badawczych, nie zaś jako funkcjonalnej (a zwłaszcza jedynej) metody ustalającej schemat postępowania. Ukryte modele Markowa są statystyczną metodą klasyfikacji sekwencji zdarzeń. Te sekwencje mogą wynikać z pewnego układu zachowania się klienta, który jest negatywnie kwalifikowany w IO lub może być wynikiem schematu przestępczego postępowania sprawcy zgodnie ze stroną przedmiotową przestępstwa. Począwszy od pierwszego $S = 1$ w dalszych etapach w HMM zwiększa się liczbę stanów o jeden tak długo, aż model osiągnie najlepsze dopasowanie.

Mając na uwadze to, że także wzrasta liczba szacowanych parametrów modelu, ważne jest również zapewnienie kryteriów informacyjnych, np. kryterium informacyjne Akaike AIC (ang. Akaike Information Criterion; Słowniczek, pkt 5). Takie działania możliwe są na potrzeby wybrania maksymalnej liczby czynników cechowania, a następnie zminimalizowania kryteriów. Ten oceniany i klasyfikowany ciąg „zdarzeń" powinien być odpowiednio parametryzowany. Parametryzacja powinna spowodować impuls do podjęcia czynności rozpoznania zachowań pod kątem „podejrzalności" w ramach systemu AML/CFT (w celu dokonania klasyfikacji do przyjętego negatywnego wzorca postępowania). Przy czym na potrzeby badawcze warte zastanowienia byłoby określenie cech jak najbardziej charakterystycznych, nie zaś ich wielościowe powiększanie dla zwiększenia trudności obliczeniowych. Możliwe jest np. przyjęcie cech, takich jak czasowość zdarzeń, liczba wykonania operacji finansowych, aktywność w jednostkach czasowych (na osi czasu), powtarzalność zachowania. Dla przykładu w zakresie AML/CFT może być to sekwencja kilku/kilkunastu podobnych transakcji dokonywanych w jednostce czasowej jako zasilających stan konta klienta lub też odwrotnie zleconych przez klienta na rzecz innego klienta (w innej lub tej samej IO) lub na rzecz różnych klientów, przez umniejszanie aktywów. Takie sytuacje były oceniane jako „surfowanie" środków w celu ich zalegalizowania lub uruchomienia kolejnych ogniw w łańcuchu dostaw dla beneficjentów terrorystycznych (tzw. skrzynka rozdzielcza, skrzynka zbiorcza). W takim przypadku

będzie można posłużyć się pewnym stanem i jego stanem następczym w najprostszej konfiguracji bez potrzeby cechowania indywidualnego, które można uzyskać na dalszym etapie procesu analizy rozpoznania AML/CFT (np. w trakcie utrzymywania stosunków gospodarczych w relacji klient – IO). Ten „ujawniony" najprostszy model podobny do wzorca dawać powinien asumpt do dalszego działania dla analityka w obszarze AML/CFT. Nie wydaje się tu niemożliwe podjęcie działań w dwóch etapach. Pierwszy mający na celu określenie jak najbardziej charakterystycznych cech. W drugim objęcie obserwacją zachowania określonego podmiotu na podstawie wykazanych uprzednio cech. Wydaje się jednak, że wobec zastosowania HMM możliwe jest podejście oparte na wzorze procesu ciągłego, ale także podejście, w którym traktowanie aktu terroru uznano by jako zestaw niezależnych zdarzeń, co można by także przetransponować na rodzaj jego finansowania. To drugie podejście powinno być rozważone, gdy zdarzenia finansowania terroryzmu (FT) są nieliczne i odmienne (od siebie, a nie od wzorca) (Krieg, Smith, Chatterjee, Chawla, 2022).

Wobec zastosowania HMM możliwe będzie wyodrębnienie trzech obszarów badawczych:

- zbadanie z zastosowaniem metody HMM sposobu aktywności organizatorskiej grupy terrorystycznej (zob. badania Vasanthan Raghavan, Aram Galstyan and Alexander G. Tartakovsky);
- zbadanie zachowania się sprawców finansowania terroryzmu opartego na produktach/usługach świadczonych przez IO (Kasianova, 2020; Aghahasanli, 2021);
- określenie zachowań wobec sprawców finansowania terroryzmu, z wykorzystaniem parametrów przyjętych dla dwóch pierwszych grup.

Jak zauważa Raul Sormani, niniejsze podejście związane jest z tym, że podstawowa motywacja modelowania działań terrorystycznych za pomocą HMM jest dwojaka. Po pierwsze, wykonywanie działalności terrorystycznej wymaga planowania i przygotowań, według etapów tworzących pewien wzór (logika przestępczego taktycznego postępowania). Ten wzór działań można modelować za pomocą łańcucha Markowa. Po drugie, terroryści pozostają wykrywalni według wskazówek na podstawie tych uskutecznianych zdarzeń w przestrzeni obserwacyjnej. Wskazówki nie są bezpośrednimi obserwacjami planowania i przygotowań, ale raczej są z nimi powiązane, co oznacza, że stany w modelu Markowa są ukryte (Sormani, 2016, s. 28). Właściwości probabilistyczne pozwalają natomiast przewidzieć sekwencję nieznanych (ukrytych) zmiennych ze zbioru obserwowanych zmiennych.

Istotne wydaje się to, przy możliwości zastosowania HMM, że w przypadku procederu FT mamy do czynienia, z punktu widzenia obserwatora, z niepełnym łańcuchem dostaw aktywów dla beneficjenta terrorystycznego (wywołane jest to zwłaszcza brakami wiedzy). Ponadto zawężone pozostaje pole dla obserwatora IO na potrzeby prowadzenia rozpoznania pod kątem uzyskania „podejrzalności" zachowań klienta i zlecanych transakcji. Stąd też wnioskowanie wstecz (retrospekcyjne)

nie może sięgać daleko, a może opierać się wyłącznie na pojedynczym fakcie (ujawnionej anomalii). Podobnie w przypadku transakcji okazjonalnej, która może okazać się jedynie jednorazowa (stąd nazwa „okazjonalna"). Tym samym decydent w instytucji obowiązanej musi działać w ograniczonym środowisku informacyjnym. HMM daje możliwości prawdopodobnego typowania przy ograniczonej liczbie zdarzeń, tak aby dalszy przewidywalny krok sprawcy był wytypowany jako zależny tylko od stanu poprzedniego i ma nie zależeć od innych stanów wcześniejszych (nieobserwowalnych lub niewidocznych dla IO). To prawdopodobieństwo nie będzie jednak typowanym jako prawdopodobieństwo zwykłego działania klienta, ale jako anomalnego określonego jako kolejny ruch w łańcuchu dostaw na rzecz beneficjenta terrorystycznego (charakterystyczne przez wypaczenie celu postępowania). A więc specyficznego dla działania sprawcy („wzorca" sprawcy przestępstwa finansowania terroryzmu). Brak przyjęcia w modelowaniu cech użytkownika wynika z potrzeby ciągłego przepływu dużych ilości danych, co wymusza inne podejście do działań predykcyjnych. Ustalenie jego „wzorca" jest możliwe z użyciem HMM. Wobec takiego podejścia każde zdarzenie wytypowane w IO jako podejrzane stać się może przyczynkiem do typowania prawdopodobieństwa kolejnego. Zwłaszcza gdy oprócz wiedzy o pojedynczym zdarzeniu obserwator nie posiada wiedzy o innych przeszłych zdarzeniach, które przyczyniły się do jego powstania, a dysponuje także zbiorem cech wytypowanych jako charakterystyczne dla ML/FT.

Tym samym HMM możliwe są do zastosowania wobec zdarzeń o znacznie ograniczonym polu informacyjnym, gdy dotyczy ono transakcji okazjonalnych czy pojedynczej aktywności klienta dotychczas „uśpionego" wobec czynnej jego roli w procederze FT. Metoda HMM pozostaje jednak na tyle uniwersalna, że możliwa jest także do zastosowania wobec anomalii związanych z oszustwami płacenia kartami kredytowymi czy identyfikacji i oceny wystąpienia zdarzeń terrorystycznych.

Ukryty model Markowa jako sposób na ujawnienie sieci sprawców ML/FT przez obserwację zachowań transakcyjnych

W omawianym przypadku niewykluczone jest w zakresie identyfikacji podmiotów-sprawców ML/FT skorzystanie z bardziej tradycyjnego podejścia grupowania klas (sekwencjonowania) nośników informacji dla zasilania (wzmacnianie) nimi agenta (jako podmiotu analizującego). Pierwsze podejście będzie dotyczyć wyodrębnienia z obsługiwanych przez IO tych transakcji, które uważane są za podejrzane. Dotyczyć to będzie także uzyskiwania tego typu informacji od innych podmiotów działających w ramach grupy czy podmiotów trzecich wobec IO. Zgromadzony materiał może być przekazywany jako zbiór uczący, a agent będzie mógł identyfikować tego typu transakcje w środowisku, a także identyfikować ich przyszłe schematy przestępcze. Podejście to dotyczy transakcji.

Drugie podejście wykorzystuje wskazany przy pierwszym materiał dla cechowania i etykietowania klienta (np. przez analizę behawioralną), który inicjuje tego typu transakcje lub jest ich beneficjentem. A tym samym uzyskanie wyniku jako profilu użytkowania zarówno w charakterystyce pojedynczego obiektu, jak i sieci powiązanych ze sobą podmiotów. W przypadku takiego podejścia można wykorzystać ukryty model Markowa, który niejako rozszyfrowuje „ukrywających" się sprawców ML/FT w całości prowadzonych transakcji w IO. Wykorzystuje się w tym zakresie własność Markowa (własność procesów stochastycznych), która polega na tym, że warunkowe rozkłady prawdopodobieństwa przyszłych stanów procesu są zdeterminowane wyłącznie przez jego bieżący stan, bez względu na przeszłość. W przypadku prania pieniędzy stan niezaobserwowany to rodzaj transakcji (podejrzanej lub normalnej).

Jak wskazano, zasadniczym problemem występującym przy budowie ukrytych modeli Markowa jest taki dobór topologii (liczba stanów i rodzaj powiązań między stanami) oraz wartości parametrów modelu, który zapewnia jego dobre działanie, np. wysoką zdolność rozpoznawania. W każdej dyskretnej chwili czasu proces znajduje się w jednym stanie oraz generowana jest obserwacja przez pewną losową funkcję. W następnej chwili czasu łańcuch Markowa przechodzi do następnego stanu zgodnie z pewnym określonym dla danego stanu prawdopodobieństwem (stan bieżącej transakcji zależy od stanu poprzedniej transakcji). Obserwator widzi tylko wynik działania losowych funkcji, przy czym nie może bezpośrednio obserwować stanów łańcucha Markowa. W swojej pracy Kseniia Kasianova przyjęła na potrzeby użycia HMM do AML określenie jako typ transakcji (podejrzana lub nie) oraz jako ukrytą (nieobserwowalną) zmienną losową, która jest zależna tylko od poprzedniej wartości (Kasianova, 2020, s. 12). Zastosowała więc HMM do zdefiniowania rodzaju transakcji każdego klienta jako albo „podejrzane", albo „normalne". Przy czym przyjęto różne cechy transakcji (np. czas transakcji, waluta, kierunki, kwota, liczba transakcji w ciągu ostatnich 7 dni itd.), które posłużyły do definiowania obserwowalnej zmiennej. Przyjęto, że zmienna obserwowalna ma dwie wartości „low-risk", „high-risk", które zdefiniowano za pomocą zmiennej pomocniczej „wynik". Zmienna „wynik" jest zmienną liczbową, która opiera się na łączeniu różnych charakterystyk transakcji, np. czy dana transakcja jest wykonywana przez podmiot ze zbioru podmiotów wyznaczonych (objętych sankcjami) lub też czy transakcja dokonywana jest z podmiotem z kraju wysokiego ryzyka. Część wyników została ustalona na podstawie reguł, które są zwykle stosowane w metodach opartych na regułach do wykrywania ML/FT, otrzymując ostatecznie określony scoring wartości. Opierając się na obserwowalnej zmiennej przewidywanej przez HMM stan ukryty z użyciem określonych prawdopodobieństw między nimi w HMM pozwalał przewidzieć sekwencję nieznane (ukryte) zmienne ze zbioru obserwowanych zmiennych. W prowadzonym badaniu wykorzystano zarówno rzeczywiste (w tym celu wykorzystano zanonimizowane dane z instytucji finansowej), jak i sztuczne dane (wygenerowane w postaci bazy danych opracowanej przez instytucję informatyczną).

Podchodząc szczegółowo do badania, należy wskazać, iż w przypadku ML stan niezaobserwowany to rodzaj transakcji (podejrzanej lub normalnej). W modelu na danych rzeczywistych celem badania wykorzystano kolejne cechy transakcji:

- data transakcji;
- kierunek transakcji (przychodząca i wychodząca);
- kwota w euro;
- waluta transakcji;
- kraj kontrahenta.

W przeprowadzonym badaniu wykorzystano także dane, które zostały wskazane przez pracowników pionu AML/CFT w zakresie klasyfikacji zdarzeń jako podejrzanych. Co istotne, w ramach badań wprowadzono świadomie dane sztuczne, tak aby wiedzieć, co badać. Zmienna „wynik" jest w tym przypadku zmienną liczbową, która opiera się na łączeniu różnych charakterystyk transakcji, np. jedną z takich cech jest sprawdzenie, czy kraj kontrahenta transakcji jest krajem wysokiego ryzyka. Im wyższa wartość zmiennej „score" (wynik), tym bardziej ryzykowna transakcja. Część wyników stanowi efekt zastosowania ustalanych reguł, które są zwykle stosowane w metodach opartych na regułach do wykrywania prania pieniędzy i zostały zbudowane na wiedzy dziedzinowej (Kasianova, 2020, s. 24). Po zdefiniowaniu wyników i zmiennych obserwacyjnych dla wszystkich badań zastosowano HMM dla każdej osoby z określonymi wartościami dla każdego przypadku. Następnie wykorzystano algorytm Bauma–Welcha do znalezienia lepszych prawdopodobieństw przejścia i emisji. Do przewidywania stanu każdej transakcji wykorzystano algorytm Viterbiego, według którego obliczano najbardziej prawdopodobny ciąg na podstawie wartości obserwowanych w kolejnych okresach. Później dla tego samego zbioru danych zastosowano algorytm grupowania k-średnich transakcji zmiennych „punktowo" każdej osoby i kolejno pogrupowano w dwa skupienia – normalne i podejrzane. Wyniki zaproponowanego modelu HMM porównano z grupowaniem k-średnich (metoda k-średnich, Słowniczek, pkt 6). Obliczono także „precyzję" i „czułość" dla HMM i k-średnich. Stwierdzono, że precyzja jest wyższa dla HMM niż dla k-średnich. Na podstawie zastosowanych metod wykazano, że HMM działa lepiej identyfikacyjnie niż algorytm grupowania k-średnich w wykrywaniu podejrzanych transakcji dla wszystkich przebadanych grup przypadków.

W podsumowaniu stwierdzono, że jednym z ograniczeń HMM jest zmienna „wynik", która opiera się na wartości ogólnego zrozumienia logiki prania pieniędzy. Im wyższa wartość zmiennej, tym bardziej ryzykowna transakcja. Co więcej, zmienna ta jest dynamiczna i należy poddawać ją przeglądowi co roku lub co kilka lat. Aby ulepszyć HMM, przydatne może być zwiększenie liczby obserwowanych poziomów zmiennych i dodanie reguł wynikających z podejścia opartego na regułach jako składnika zmiennych wyniku (Kasianova, 2020, s. 41). W badaniu HMM określiło 86% wszystkich osób jako podmioty normalne i rzeczywiście pozostawało to zgodne ze stanami niepodejrzanymi. Otrzymano również 8,1% wszystkich transakcji

jako podejrzanych według modelu, ale oznaczone jako normalne w zbiorze danych. Transakcje te powinny zostać – na dalszym etapie analizy rozpoznania – zbadane przez specjalistę podmiotu AML, aby upewnić się, że naprawdę nie są podejrzane. W badaniu przeprowadzonym wśród osób, które posiadały tylko normalne transakcje, 6,4% transakcji zostało oznaczonych jako podejrzane przez HMM, dlatego także one powinny zostać sprawdzone przez specjalistę ds. przeciwdziałania praniu pieniędzy.

Jako uzupełnienie przedmiotowego podejścia można przedstawić także wyniki prac Ismayila Aghahasanliego (2021). Według ustalonego stanu ukrytego generowane są obserwowalne wartości zmiennych, czyli kwota transakcji, waluta, czas transakcji, rodzaj kontrahenta. Wzięto także pod uwagę kilka zmiennych pomocniczych, takich jak suma całkowitej kwoty transakcji w ciągu jednego dnia i liczba transakcji w ciągu jednego dnia dla każdego klienta banku w momencie realizacji transakcji. HMM posłużyło do przewidywania stanu ukrytego za pomocą obserwowalnych zmiennych. W modelu opartym na zmiennych obserwowalnych zastosowano pojedynczą zmienną obserwowalną skonstruowaną w celu sklasyfikowania transakcji jako: niskiego, średniego lub wysokiego ryzyka. Ważnym założeniem było również to, że instytucja finansowa może wykryć jedynie podejrzane działania swoich klientów, a nie faktyczny zamiar prania pieniędzy lub nielegalne zachowanie (Aghahasanli, 2021, s. 7). Istotne stało się początkowo skupienie na określeniu modelu normalnego zachowania klienta, tak aby można było ustalić zachowania anomalne.

Do analizy transakcji posłużono się określonymi zmiennymi:

- user_id – jednoznacznie definiuje osobę, którą jest kontrahent transakcji w ramach banku;
- typ – pokazuje kierunek transakcji (wychodząca lub przychodząca);
- date_completed – pokazuje datę i godzinę przeprowadzenia/zakończenia transakcji;
- from_cur i to_cur – pokazują walutę transakcji;
- kwota_w_eur – oznacza kwotę transakcji w ekwiwalencie w euro;
- meta_sar_id – identyfikuje, czy transakcja została wygenerowana celowo jako podejrzana, a numer identyfikacyjny podaje dokładny scenariusz. Posłużono się oznaczeniem binarnym „0", gdy transakcja jest normalna (prawidłowa), oraz „1", gdy transakcja jest podejrzana (anomalna);
- kraj_kontrahent – pokazuje kraj pochodzenia kontrahenta transakcji.

Główną ideą jest to, że transakcje są uważane za normalne lub podejrzane i fakt ten jest traktowany jako stany ukryte w HMM.

Następnym krokiem w zastosowaniu HMM pozostawało zdefiniowanie zmiennej „obserwowalne". We wszystkich trzech przeprowadzonych badaniach zmienną pośrednią utworzono z kombinacji liniowej cech transakcyjnych, które są uważane za czynniki ryzyka w walce z praniem pieniędzy. Kompozycja ta została zaczerpnięta z praktyki AML (Aghahasanli, 2021, s. 30). Składniki „wyniku" są pobierane przez reguły, które są zwykle używane w metodzie opartej na regułach do wykrywania

prania pieniędzy i zostały zbudowane z użyciem wiedzy dziedzinowej. Jeżeli wynik przekroczyłby określony próg, to w zależności od danej jest on kwalifikowany jako: wysokiego, średniego czy niskiego ryzyka. Im wyższa wartość zmiennej „score", tym bardziej ryzykowna jest transakcja. Jednocześnie utworzone zostały dodatkowe zmienne:

- sum_1in – suma kwot transakcji przychodzących w ciągu 1 (jednego) dnia dla tego samego użytkownika;
- sum_1out – suma kwot transakcji wychodzących w ciągu 1 (jednego) dnia dla tego samego użytkownika;
- count_1 – liczba transakcji w ciągu 1 (jednego) dnia dla tego samego użytkownika.

Po przeanalizowaniu danych zmienna „wynik" dla przypadku badawczego 1 i 2 została wyznaczona ręcznie na podstawie przyjętych zasad (np. oparte na czasie transakcji i wysokości kwoty transakcji). Po oszacowaniu zmiennej score dla każdej transakcji maksymalna zmienna score została obliczona dla każdego użytkownika. Ostatnim krokiem było zastosowanie HMM do zdefiniowania stanu ukrytego na podstawie sekwencji obserwacji. Prawdopodobieństwa przejścia i emisji obliczono, stosując algorytm Bauma–Welcha (Kwok, 2019; Słowniczek, pkt 1), a następnie do oszacowania stanu „ukrytego" wykorzystano algorytm Viterbiego (Wikipedia, 2024; Słowniczek, pkt 2) dla każdej transakcji. We wszystkich przypadkach zastosowano te same algorytmy. W podsumowaniu przedstawiono, że HMM wskazał tylko na te transakcje jako podejrzane, które mają wysokie ryzyko w zmiennej obserwacyjnej, w tym zidentyfikowane jako fałszywie pozytywnie. Tym samym uznano, że wykrywanie stanów ukrytych jest silniejsze od k-średnich (Słowniczek, pkt 6) i zależy od rodzaju obserwowanej zmiennej. Żadna z transakcji niskiego ryzyka nie okazała się podejrzana, a 0,08% znalazło się w kategorii wysokiego ryzyka na 0,1% faktycznie podejrzanych transakcji (badanie pierwsze) (Aghahasanli, 2021, s. 36).

W drugim badaniu przeanalizowano 6854 transakcji, które uznano za podejrzane. HMM poprawnie przewidział 5771 (84%) z nich jako podejrzane. Ponieważ w tym badaniu wszystkie transakcje były podejrzane, precyzja dla obu modeli mieściła się w maksimum wyniku. HMM błędnie przewidział jako normalne tylko 16% transakcji. W trzecim badaniu ogółem przeanalizowano 55275 transakcji i 32 użytkowników. Podejrzanych było 235 transakcji i 11 użytkowników. HMM naprawdę przewidział 95 (40%) transakcji i 9 użytkowników jako podejrzanych. HMM błędnie zidentyfikował 83 transakcje i 5 użytkowników jako podejrzanych, którzy nimi nie byli. Uzyskane rezultaty związane były ze zdefiniowaniem zmiennej „wynik" i przyjętych czynników wyróżniających, co miało pozytywne przełożenia na identyfikacje transakcji wysokiego ryzyka i uznania ich w zakresie kategorii „podejrzalności". Niemniej jednak jako wniosek wskazano, że wygenerowane w określony sposób efekty wymagają jeszcze dodatkowo na potrzeby oceny ostatecznej przeprowadzenia przez eksperta oceny komórki AML/CFT (Aghahasanli, 2021, s. 40).

Thayne R. Coffman i Sherry E. Marcus również badali powiązania, tym razem występujące pomiędzy podmiotami terrorystycznymi, z wykorzystaniem metody HMM (Coffman, Marcus, 2004). Przyjęta przez nich analiza sieci społecznościowych (SNA, ang. Social Network Analysis) jako model analityczny przedstawia komunikację interpersonalną w postaci skierowanych grafów. Tym samym w ramach zastosowania metody HMM można uznać, że przedmiotem rozważań jest pewien graf skierowany, w którym wagi na przejściach oznaczają prawdopodobieństwo przejścia do stanu następnego. Metryki SNA określają ilościowo różne aspekty wzorców komunikacji w grupie. Celem pracy była identyfikacja komunikatów terrorystycznych na podstawie ich nietypowych wartości metryki SNA. Strukturę społeczną grup terrorystycznych i innych nielegalnych organizacji odróżniano od normalnych grup tym, że ich wartości metryczne zmieniają się w różny sposób w czasie. W tym przypadku wykorzystano ukryte modele Markowa, aby identyfikować grupy wykazujące podejrzane ewolucje. Wykorzystano całą historię struktury społecznej, a nie tylko oglądanie struktury w jednym momencie. W przyjętym studium przypadku autorzy osiągnęli 96% dokładność klasyfikacji na nowych danych syntetycznych z użyciem dwóch 35-stanowych jednowymiarowych HMM przeszkolonych do modelowania normalnych i podejrzanych ewolucji metryki charakterystycznej długości ścieżki.

Tozammel Hossain, Shuyang Gao, Brendan Kennedy, Aram Galstyan i Prem Natarajan (2020) także zaproponowali zastosowanie modelu HMM na potrzeby prowadzenia rozpoznania zagrożenia terroryzmem i gwałtownych zdarzeń militarnych określonych jako MANSA. Badanie posłużyło do określenia występowania zdarzeń gwałtownych w obszarze regionu Bliskiego Wschodu i Afryki Północnej (MENA). Jako źródła informacji wytypowano: sprawców (ISIS i Syryjska Armia Arabska), cel działania, czas i lokalizację oraz dane uzyskane z artykułów, jakie w okresie badawczym ukazywały się w mediach. Podstawowym założeniem dla wykorzystania HMM była teza, że aktualna liczba zdarzeń (np. działań o charakterze terrorystycznym) zależy od przeszłej historii wydarzeń za pośrednictwem K dominujących państw ukrytych, które reprezentują różne fazy operacyjne działalności terrorystycznej. Proces przechodzi probabilistycznie pomiędzy stanami niskiej H i wysokiej aktywności L. Będąc w konkretnym stanie proces generuje pewne zdarzenia zgodnie z rozkładem prawdopodobieństwa zależnym od stanu poprzedniego (wzięto pod uwagę główne dwa stany wysokiej i niskiej aktywności). Stany ukryte Z mają wartość dyskretną zmienną losowo – ukryty stan charakteryzuje określony tryb operacyjny organizacji. Przejście między stanami jest przejściem Markowa, czyli stan przyszły jest warunkowo niezależny od stanów przeszłych, biorąc pod uwagę jednak stan bieżący. Jednocześnie należałoby zwrócić uwagę na kwestię doboru do badania danych zewnętrznych, zwłaszcza tych wyselekcjonowanych z mediów społecznościowych czy Internetu. W takim przypadku należy mieć na uwadze ustalenie i ujawnienie słów kluczowych ujawnionych w tekstach, generując je z treści przeglądanych mediów (w celu uzyskania dla dalszych badań danych demograficznych, przestrzennych czy czasowych).

Ma to istotne znaczenie dla nauczenia się modelu procesu, który wytworzył te zdarzenia. Model ten, w dalszej kolejności, może być używany do przewidywania nowych zdarzeń. Dlatego też ze źródeł zewnętrznych potrzebne jest identyfikowanie prekursorów dla przyszłych zdarzeń.

Nie sprawdziło się podejście badaczy opierających się jedynie na gromadzeniu danych historycznych i braku informacji w czasie rzeczywistym (Porter, White, 2012). Stąd w tym badaniu wykorzystano modele, proponując użycie dodatkowych (zastępczych) źródeł danych/zewnętrznych, aby zrekompensować brak najnowszych danych o zdarzeniach (inne dokumenty zawierające sygnaturę czasową i odnoszące się do zdarzeń terrorystycznych, stanowiące stan rozbudowanego kontekstu zdarzeń). W prowadzonym badaniu korzystano pierwotnie z raportów analitycznych wykonywanych ręcznie przez instytucję analityczną, co powodowało potrzebę oczyszczenia danych z błędów, np. stosowania duplikatów. Niemniej stosowanymi atrybutami były następujące dane: sprawca, status sprawcy, przybliżona lokalizacja, przyczyny, kraj, najwcześniejsza zgłoszona data zdarzenia, komentarz kodowania, data zdarzenia, identyfikator zdarzenia, podtyp zdarzenia, typ zdarzenia, pierwszy zgłoszony link, szerokość geograficzna, długość geograficzna, źródło wiadomości, inne linki, data rewizji, stan, cel, nazwa celu i status celu. Przy rozwiązywaniu problemów brano pod uwagę prawdopodobieństwa emisji zdarzenia jako wartość ciągłą spośród czterech możliwych rozkładów: Poissona, Gaussa, geometrycznego lub modelu płotkowego. Przy czym liczba czynności wykonywanych przez sprawcę może mieć ukryte struktury (np. okresy wysokiej i niskiej aktywności), które mogą nie być dobrze uchwycone przy użyciu prostego procesu liczenia, takiego jak proces Poissona. Stąd też wykorzystano HMM do przechwytywania ukrytych struktur w działaniach różnych podmiotów w danym zbiorze. Zbiory danych zawierały dzienne liczby zdarzeń terrorystycznych ISIS w Iraku i Syrii w latach 2016-2017.

W ramach badania modelu wskazano we wnioskach na to, że sprawca jest bardziej skłonny do pozostania w tym samym stanie ukrytym niż przejście do nowego. Ponadto, co być może ważniejsze, tempo zdarzeń scharakteryzowane przez średnią zdarzeń (scharakteryzowane przez średnią modelu gaussowskiego) znacznie różni się między stanami. Średnia liczba ataków dziennie wynosi 13:4, gdy sprawca znajduje się w stanie wysokiej aktywności H, w porównaniu do 5:7, gdy jest w stanie niskiej aktywności L (Hossain, Gao, Kennedy et al., 2020). W zakresie ogólnych spostrzeżeń odnoszących się do przedmiotowego badania należałoby zwrócić uwagę na to, że zastosowane metody oparte na stanie (HMM) i wprowadzone autoregresyjne modele (jako modele statystyczne używane do przewidywania przyszłych danych na podstawie przeszłych wartości w szeregach czasowych), które posłużyły do generowania prognoz zdarzeń wraz ze wskaźnikami zewnętrznymi (np. do generowania cech czasowych z ciągów tekstowych w źródłach medialnych). Modele te zakładają, że podstawowy proces generowania danych jest liniowy, to znaczy, że wartość w punkcie czasowym jest liniową kombinacją wartości z przeszłości.

Jednak rzeczywiste szeregi czasowe wykazują zmienność i nieliniowość. Zarówno modele HMM, jak i RARE radzą sobie całkiem dobrze przy rozsądnej ilości danych (zdarzenia na poziomie aktora i kraju), natomiast wydajność pogarsza się, gdy zdarzenia są rzadkie. Gdy gęstość zdarzeń jest niska, a typ zdarzeń rzadki, stanowi to wyzwanie dla proponowanych modeli przewidywania zdarzeń w takich przypadkach ustawienia. Przy niskim poziomie gęstości zdarzeń modele HMM i autoregresji wydają się być niewystarczające. Dla rzadkich zdarzeń, HMM i modele autoregresyjne mogą nie działać dobrze. Aby rozwiązać te problemy, niezbędne byłoby posiadanie modeli predykcyjnych, które uwzględniłyby skomplikowany kontekst zdarzenia oraz brały pod uwagę źródła zewnętrzne. Niemniej autorzy (Hossain, Gao, Kennedy et al., 2020) przewidują, że zaproponowany model będzie mógł czerpać także informacje zewnętrzne (np. których źródłem są: artykuły prasowe, Twitter, blogi), co mogłoby polepszyć uzyskanie precyzyjniejszych wyników badań.

Podsumowanie

HMM jest uniwersalnym podejściem z zakresu statystyki analitycznej o charakterze probabilistycznym. Dotychczasowe badania w zakresie zastosowania HMM do przewidywania zagrożeń związanych z ML/TR wykazały z jednej strony pewne trudności wobec kreowania zbiorów danych uczących celem zastosowania modelu, z drugiej zapewniły rezultaty potwierdzające sprawdzalność samego modelu HMM wobec typowania podejrzanych transakcji lub kreowania zdarzeń o charakterze terrorystycznym. Dzięki zastosowaniu HMM możliwe jest typowanie probabilistyczne kolejnych kroków sprawców, zwłaszcza przy ograniczonej wiedzy dotyczącej samego zdarzenia. Pomocniczym staje się tu nie tylko właściwe cechowanie informacji do zbiorów uczących, typowanie elementów ukrytych łańcucha przejść, lecz także korzystanie z innych zewnętrznych źródeł informacji zawierających identyfikacyjne prekursory zdarzeń.

SŁOWNICZEK

[1] ALGORYTM BAUMA–WELCHA – znany również jako algorytm do przodu i do tyłu, jest podejściem do programowania dynamicznego i szczególnym przypadkiem algorytmu maksymalizacji oczekiwań (algorytm EM). Jego celem jest dostrojenie parametrów HMM, czyli macierzy przejść stanów A, macierzy emisji B i rozkładu stanu początkowego π_0, tak aby model maksymalnie przypominał obserwowane dane. Algorytm Bauma–Welcha polega na znalezieniu w kolejnych iteracjach lokalnego maksimum funkcji wiarygodności.

[2] ALGORYTM VITERBIEGO – algorytm dekodujący, o strategii programowania dynamicznego, opracowany przez Andrew Viterbiego i opublikowany przez niego w 1967 roku w IEEE Transactions on Information Theory, IT-13 w artykule *Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm* (Wikipedia, 2024).

[3] DYNAMICZNA OCENA RYZYKA – jest to proaktywne podejście do zarządzania ryzykiem, polegające na ciągłej ewaluacji i ponownej ocenie potencjalnych zagrożeń w miarę zmieniających się okoliczności.

[4] INFORMATYKA TERRORYSTYCZNA – została zdefiniowana jako zastosowanie zaawansowanych metodologii, technik łączenia i analizy informacji w celu pozyskiwania, integrowania przetwarzania, analizowania i zarządzania różnorodnością informacji związanych z terroryzmem dla zastosowań związanych z bezpieczeństwem międzynarodowym i wewnętrznym. Różnorodność metod stosowanych w informatyce terrorystycznej pochodzi z informatyki, statystyki, matematyki, lingwistyki, nauk społecznych i polityki publicznej. Obejmują one gromadzenie masy informacji z wielu źródeł i w wielu językach (Chen, Reid, Sinai et al., 2008).

[5] KRYTERIUM INFORMACYJNE AKAIKE AIC – służy do próby zmierzenia względnej jakości modeli ekonometrycznych objaśniających – zmienną zależną dla danego zbioru danych, AIC zapewnia oszacowanie informacji, które zostałyby utracone, gdyby konkretny model został wykorzystany do realizacji procesu, który wytworzył dane. Celem wyboru modelu według AIC jest oszacowanie straty informacji, gdy rozkład prawdopodobieństwa f związany z prawdziwym (generującym) modelem jest aproksymowany rozkładem prawdopodobieństwa g związanym z modelem, który ma być oszacowany (Piłatowska, 2009).

[6] METODA K-ŚREDNICH – metoda należąca do grupy algorytmów analizy skupień, tj. analizy polegającej na szukaniu i wyodrębnianiu grup obiektów podobnych (skupień). Reprezentuje ona grupę algorytmów niehierarchicznych. Główną różnicą pomiędzy niehierarchicznymi i hierarchicznymi algorytmami jest konieczność wcześniejszego podania ilości skupień. Za pomocą metody k-średnich zostanie utworzonych k różnych możliwie odmiennych skupień. Algorytm ten polega na przenoszeniu obiektów ze skupienia do skupienia tak długo, aż zostaną zoptymalizowane zmienności wewnątrz skupień oraz pomiędzy skupieniami. Oczywiste jest, iż podobieństwo w skupieniu powinno być jak największe, osobne skupienia zaś powinny się maksymalnie od siebie różnić (Kajstura, 2024).

[7] MODELE AUTOREGRESYJNE – rodzaj modeli statystycznych stosowanych w analizie szeregów czasowych, które przyjmują serię z góry określonych wartości i przewidują przyszłe wartości na podstawie wartości z przeszłości. Podstawowym założeniem tych modeli jest to, że seria wartości zaobserwowanych w przeszłości pozostanie w pewnym stopniu stała w najbliższej przyszłości. Modele AR są szczególnie przydatne do przewidywania trendów w danych, które zmieniają się sezonowo lub wykazują zachowanie cykliczne (FineProxy, 2024).

[8] PROCES STOCHASTYCZNY – ogólnie określa się jako funkcję $y(t)$ zależną od parametru czasowego t, której wartości w każdej chwili są zmiennymi losowymi; jeżeli przez parametr t może przebiegać jedynie zbiór dyskretny (np. obserwacje są dokonywane tylko w ustalonych chwilach), to proces taki redukuje się do ciągu zmiennych losowych (Encyklopedia PWN, 2024).

[9] ŚRODKI BEZPIECZEŃSTWA FINANSOWEGO – zgodnie z art. 34 ust. 1 ustawy o p.p.p.f.t. środki bezpieczeństwa finansowego obejmują m.in. ocenę stosunków gospodarczych i uzyskanie informacji stosownie do sytuacji na temat ich celu i zamierzonego charakteru oraz bieżące monitorowanie stosunków gospodarczych klienta (GIIF, 2024).

BIBLIOGRAFIA

[1] Aghahasanli, I., 2021. *Detecting Money Laundering in Transaction Monitoring Using Hidden Markov Model*, Tartu: University of Tartu, Faculty of Science and Technology Institute of Mathematics and Statistics, https://dspace.ut.ee/bitstream/handle/10062/72862/aghahasanli_ismayil_msc_2021.pdf?sequence=1&isAllowed=y (dostęp: 4.05.2024).

[2] BusinessInsider, 2021. *Model Markowa – co to jest i jakie ma zastosowanie?*, https://businessinsider.com.pl/gospodarka/makroekonomia/model-markowa-definicja-typy-i-zastosowanie/6czkt2x (dostęp. 10.02.2024).

[3] Chen, H., Reid, E., Sinai, J., Silke, A., Ganor, B. (red.), 2008. *Terrorism Informatics. Knowledge Management and Data Mining for Homeland Security*, New York: Springer.

[4] Coffman, T.R., Marcus, S.E., 2004. *Dynamic Classification of Groups Through Social Network Analysis and HMMs*, Aerospace Conference 2004, *Proceedings. 2004 IEEE*, vol. 5, s. 3197-3205.

[5] Encyklopedia PWN, 2024. *Stochastyczne procesy*, https://encyklopedia.pwn.pl/haslo/stochastyczne-procesy;3979832.html (dostęp: 11.04.2024).

[6] Fabian, P., Szedel, J., 2002. Teoretyczne podstawy zastosowań ukrytego modelu Markowa do rozpoznawania wzorców, *Studia Informatica*, vol. 23, nr 4(51), s. 301-328.

[7] Figielska, E., 2011. Ewolucyjne metody uczenia ukrytych modeli Markowa, *Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki*, nr 5, s. 63-74.

[8] Filipowicz, B., 1996. *Modele stochastyczne w badaniach operacyjnych*, Warszawa: Wydawnictwa Naukowo-Techniczne.

[9] FineProxy, 2024. *Modele autoregresyjne*, https://fineproxy.org/pl/wiki/auto-regressive-models/ (dostęp: 8.03.2024)

[10] Genge, E., 2014. Zastosowanie ukrytych modeli Markowa w analizie oszczędności wśród Polaków, *Zeszyty Naukowe Wydziałowe Uniwersytetu Ekonomicznego w Katowicach. Studia Ekonomiczne*, nr 189, s. 58-66.

[11] GIIF, 2024. Komunikat nr 22 w sprawie praktycznych aspektów stosowania środków bezpieczeństwa finansowego oraz przekazywania zawiadomień, o których mowa w art. 74 i art. 86 ustawy AML, https://www.gov.pl/web/finanse/komunikat-nr-22-w-sprawie-praktycznych-aspektow-stosowania-srodkow-bezpieczenstwa-finansowego-oraz-przekazywania-zawiadomien-o-ktorych-mowa-w-art-74-i-art-86-ustawy-aml2 (dostęp: 20.05.2024).

[12] Hossain, T., Gao, S., Kennedy, B., Galstyan, A., Natarajan, P., 2020. Forecasting Violent Events in the Middle East and North Africa Using the Hidden Markov Model and Regularized Autoregressive Models, *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, vol. 17, nr 3, s. 269-283.

[13] Kasianova, K., 2020. *Detecting Money Laundering Using Hidden Markov Model*, Tartu: University of Tartu, Faculty of Social Sciences, School of Economics and Business Administration, https://dspace.ut.ee/bitstream/handle/10062/68089/kasianova_kseniia.pdf?sequence=1&isAllowed=y (dostęp: 17.02.2024).

[14] Kajstura, A., 2024. *Metoda k-średnich*, https://www.statystyka.eu/analiza-skupien/metoda-k--srednich.php (dostęp: 2.03.2024).

[15] Krieg, S.J., Smith, C.W., Chatterjee, R., Chawla, N.V., 2022. Predicting Terrorist Attacks in the United States Using Localized News Data, *PLoS ONE*, vol. 17, nr 6.

[16] Kwok, R., 2019. *Baum-Welch algorithm for training a Hidden Markov Model – Part 2 of the HMM series*, https://medium.com/analytics-vidhya/baum-welch-algorithm-for-training-a-hidden-markov-model-part-2-of-the-hmm-series-d0e393b4fb86 (dostęp: 20.05.2024).

[17] Łochowski, R.M., 2019. *Modele Markowa i analiza przeżycia w ubezpieczeniach*, Szkoła Główna Handlowa w Warszawie, https://web.sgh.waw.pl/~rlocho/wyk5_MMSA.pdf (dostęp: 16.03.2024).

[18] Mazurek, M., 2010. Ukryte modele Markowa jako metoda eksploracji danych tekstowych, *Biuletyn Instytutu Systemów Informatycznych*, nr 6, s. 27-31.

[19] Pattipati, K., Willett, P., Allanach, J., Tu, H., Singh, S., 2006. *Hidden Markov Models and Bayesian Networks for Counter-Terrorism*, [w:] Popp, R., Yen, J. (red.), *Emergent Information Technologies and Enabling Policies for Counter Terrorism*, New York: Wiley-IEEE Press, s. 27-50.

[20] Piłatowska, M., 2009. Prognozy kombinowane z wykorzystaniem wag Akaike'a, *Acta Universitatis Nicolai Copernici. Ekonomia*, nr 39.

[21] Porter, M., White, G., 2012. Self-Exciting Hurdle Models for Terrorist Activity, *The Annals of Applied Statistics*, vol. 6, nr 1, s. 106-124.

[22] Przybysz, P., Kasprzak, W., 2012. *Rozpoznawanie zdań w sygnale mowy z wykorzystaniem modelu HMM*, Raport IAiIS PW, nr 12-05, Warszawa: Politechnika Warszawska, Wydział Elektroniki i Technik Informacyjnych, Instytut Automatyki i Informatyki Stosowanej, https://www.ia.pw.edu.pl/~wkasprza/PAP/Raport_RozMo_12-05.pdf (dostęp: 7.03.2024).

[23] Raghavan, V., Galstyan, A., Tartakovsky, A.G., 2013. Hidden Markov Models for the Activity Profile of Terrorist Groups, *The Annals of Applied Statistics*, vol. 7, nr 4, s. 2402-2430.

[24] Sormani, R., 2016. *Criticality Assessment of Terrorism Related Events at Different Time Scales*, https://boa.unimib.it/retrieve/e39773b3-3524-35a3-e053-3a05fe0aac26/phd_unimib_075030.pdf (dostęp: 17.05.2024).

[25] Ustawa, 2018. Ustawa z dnia 1 marca 2018 r. o przeciwdziałaniu praniu pieniędzy oraz finansowaniu terroryzmu (Dz.U. 2023 poz. 1124, 1285, 1723, 1843; Dz.U. 2024 poz. 850).

[26] Wikipedia, 2024. *Algorytm Viterbiego*, https://pl.wikipedia.org/wiki/Algorytm_Viterbiego (dostęp: 20.05.2024).