Państwo i społeczeństwo

Analiza danych z faktury ustrukturyzowanej

Możliwości wykorzystania uczenia maszynowego w KSeF

W ostatniej dekadzie obserwujemy dynamiczny rozwój technologii sztucznej inteligencji (ang. *artificial intelligence* – AI), w szczególności jej podzbioru – uczenia maszynowego (ang. *machine learning* – ML), skupionego na algorytmach uczących się na podstawie danych. Od kilkunastu lat w Polsce widoczna jest zmiana podejścia do przekazywania informacji organom podatkowym, od raportowania w formie standardowej do raportowania w czasie rzeczywistym. Ta cecha powoduje, że dane z faktury ustrukturyzowanej mogą być w sposób efektywny i skuteczny wykorzystane w ramach uczenia maszynowego. Zastosowanie AI w obszarze faktury ustrukturyzowanej oznacza przede wszystkim automatyzację procesów księgowych, analizę danych w czasie rzeczywistym oraz zwiększenie efektywności kontroli i czynności sprawdzających. Sztuczna inteligencja może wspierać zarówno firmy w poprawnym generowaniu danych, jak i organy podatkowe w wykrywaniu nieprawidłowości. Artykuł ma na celu zarysowanie obecnego stanu wiedzy na temat zastosowań AI w analizie danych gospodarczych oraz wskazanie kierunków dalszych badań w związku z rozszerzeniem wolumenu danych o informacje z faktury ustrukturyzowanej.

 DOI: 10.53122/ISSN.0452-5027/2026.1.28

LESZEK SYGUŁA

Uczenie maszynowe

Doświadczenia związane z wykorzystywaniem systemów opartych na wiedzy zakodowanej na stałe sprawiają, że od agentów sztucznej inteligencji oczekuje się zdolności do samodzielnego zdobywania wiedzy przez ekstrakcję wzorców z surowych danych. Zdolność ta jest znana jako uczenie maszynowe[1]. Uczenie maszynowe to nauka o algorytmach i systemach poprawiających swoją wiedzę oraz wyniki wraz ze zdobywanym doświadczeniem[2]. Jest podzbiorem sztucznej inteligencji, która na podstawie wyszukiwania relacji w danych uczących próbuje naśladować inteligentne zachowania (np. wyszukiwanie anomalii).

Jedną z dziedzin najszybciej wdrażających narzędzia sztucznej inteligencji są finanse. Rozwiązania takie, jak uczenie maszynowe (nadzorowane, nienadzorowane, wzmacniane) dostarczają odpowiedzi na wyzwania dotyczące np. analizy dużych zbiorów danych i wyciągania wniosków.

Obecnie wykorzystujemy w finansach znaczną liczbę informacji z systemów takich jak: Jednolity Plik Kontrolny (JPK), System Teleinformatyczny Izby Rozliczeniowej (STIR), System Elektronicznego Nadzoru Towarów (SENT), System informacji o stosowanych schematach podatkowych (MDR). Od sposobu ich łącznego użytkowania zależy większa efektywność finansowa, zachowanie konkurencyjności na rynku, uszczelnienie systemu podatkowego i zwiększenie wpływów do budżetu państwa. Analizowanie zależności występujących w danych stało się więc istotnym elementem życia gospodarczego.

W ostatnich latach obserwuje się wzrost zainteresowania naukowców możliwościami zastosowania w finansach sztucznej inteligencji i uczenia maszynowego. Jak wskazują K. Prędkiewicz i K. Biegun, S. Aziz i in. dokonali przeglądu literatury w tym zakresie przy użyciu podejścia bibliometrycznego – przeanalizowali łącznie 5942 artykuły naukowe od 1990 r. do 2020 r.[3] Zidentyfikowali następujące nurty badawcze w analizowanym obszarze, tj.:

- przewidywanie niewypłacalności (w wypadku dużych zbiorów danych ML okazuje się skuteczniejsze niż metody eksperckie)[4],
- przewidywanie cen akcji i zarządzanie portfelem,
- przewidywanie cen ropy[5],
- przeciwdziałanie praniu brudnych pieniędzy[6],
- finanse behawioralne.

---

1 I. Goodfellow, Y. Bengio, A. Courville: *Deep Learning*. MIT Press, Cambridge 2016; <http://www.deeplearningbook.org>.

2 M. Wróbel, P. Wojda: *Możliwości optymalizacji procesu transportowego z wykorzystaniem sztucznej inteligencji*, „Journal of TransLogistics" nr 14/2018, s. 197-204.

3 K. Prędkiewicz, K. Biegun: *AI – sztuczna inteligencja w finansach przedsiębiorstw* [w:] *Finanse przedsiębiorstw*, M. Sierpińska i P. Szczepanowski (red.), Polska Akademia Nauk 2024, s. 319-332.

4 A. Paliński: *Metody uczenia maszynowego w prognozowaniu niewypłacalności*, „Studia Ekonomiczne" nr 358/2018, s. 173-181.

5 W. Sikorski: *Sztuczna inteligencja, uczenie maszynowe i analiza big data w energetyce przyszłości*, „Kierunek Energetyka" nr 3/2024.

6 M. A. Kędzierski: *Propozycja wykorzystania uczenia przez wzmacnianie w celu optymalizowania podejmowania decyzji w zakresie przeciwdziałania praniu pieniędzy oraz finansowania terroryzmu (część 2)*, „Nowoczesne Systemy Zarządzania" nr 4/2023, s. 49-68.

Innymi przykładami zastosowania uczenia maszynowego w obszarze finansów są:
1. Automatyzacja sprawozdawczości finansowej i analizy. Systemy oparte na uczeniu maszynowym automatyzują zbieranie, analizowanie i przedstawianie danych finansowych, w szczególności konieczność przetworzenia znacznej liczby danych „z niespotykaną dotąd szybkością, dostarczając w czasie rzeczywistym informacji oraz prognoz, które są kluczowe dla strategicznego podejmowania decyzji".
2. Wykrywanie oszustw i zarządzanie ryzykiem. Analiza wzorców i anomalii w danych transakcyjnych pozwala na identyfikowanie potencjalnych oszustw i ryzyka finansowego.
3. Integracja i optymalizacja łańcucha przepływów finansowych (Financial Supply Chain Management, FSCM). Zastosowanie uczenia maszynowego optymalizuje proces zarządzania przepływem kapitału, może zredukować koszty operacyjne. Umożliwia „analizę danych z wielu źródeł, przewidywanie opóźnień w płatnościach, optymalizację terminów płatności i dostaw itd. Przez zastosowanie algorytmów przewidywania i optymalizacji firmy mogą lepiej zarządzać swoim kapitałem obrotowym, minimalizować ryzyko kredytowe i poprawiać płynność finansową.
4. Automatyzacja i usprawnienie procesów rozliczania podatków. Uczenie maszynowe przyczynia się również do optymalizacji procesów składania informacji podatkowych. Dzięki niemu „można automatycznie gromadzić i analizować dane finansowe, wypełniać formularze podatkowe oraz identyfikować potencjalne oszczędności podatkowe lub błędy, które mogłyby zostać przeoczone przez ludzi. Automatyzacja ta może znacznie zmniejszyć koszt przygotowania i składania zeznań podatkowych, a także zwiększyć dokładność i zgodność z przepisami podatkowymi, co jest kluczowe dla zapobiegania sankcjom"[7].

W procesie uczenia maszynowego nadzorowanego stosuje się zbiór danych ze zmienną prognozowaną, zawierającą przypisane już wcześniej etykiety (ang. *labels*). Dzięki nim algorytm uczy się wzorców, czyli tego, jak poprawnie przypisywać etykiety lub wartości w przyszłości, w sytuacji gdy dysponuje się jedynie zbiorem cech niezależnych (np. ceną, czasem trwania, liczbą produktów). Modele o wysokiej skuteczności zazwyczaj wdraża się produkcyjnie. Pozostałe są poprawiane aż do momentu osiągnięcia przez nie wysokiej trafności typowań.

Do najczęściej stosowanych algorytmów uczenia nadzorowanego można zaliczyć m.in.: regresję liniową, wielomianową oraz logistyczną, drzewa decyzyjne i lasy losowe, maszynę wektorów nośnych (ang. *support vector machine*, SVM), sztuczne sieci neuronowe. Kolejna metoda to uczenie nienadzorowane, w którym dane uczące nie posiadają etykiet. W tym modelu szuka się powiązań między danymi, wykorzystując algorytmy analizy skupień (ang. *clustering*), to jest klasteryzację albo grupowanie. Przykładem problemu analizy skupień jest określenie grup podobnych użytkowników, np. prawidłowo

[7] Wybrane przykłady przedstawione za K. Prędkiewicz, K. Biegun: *AI – sztuczna inteligencja*…, op.cit, s. 323.

rozliczających się z obowiązków podatkowych lub dokonujących oszustw. Uczenie nadzorowane i uczenie nienadzorowane są jednymi z głównych metod uczenia wykorzystywanych w uczeniu maszynowym. Wyróżnić można jeszcze uczenie wzmacniane. W przeciwieństwie do uczenia nadzorowanego i nienadzorowanego w tym modelu nie przygotowuje się zestawu danych uczących, tylko środowisko, z którego model będzie zbierał dane automatycznie. Podejście to jest stosowane, gdy najlepsza decyzja zależy od sekwencji działań, a nie od pojedynczej prognozy.

Krajowy System e-Faktur (KSeF), jako zbiór danych Szefa Krajowej Administracji Skarbowej, jest zbiorem typu Big Data. Wobec powyższego będą na nim prowadzone typowe analizy Big Data dotyczące eksploracji danych, analizy biznesowej, diagnostyki i raportowania. Zastosowanie metod uczenia maszynowego pozwoli natomiast na wykorzystanie zbioru danych KSeF do wychwytywania wzorców, przewidywań i podejmowania decyzji[8].

Krajowy System e-Faktur

Krajowy System e-Faktur – ogólnokrajowa platforma informatyczna do wystawiania, przesyłania, otrzymywania i przechowywania faktur ustrukturyzowanych (w formacie XML) – został wprowadzony do porządku prawnego ustawą z 29 października 2021 r. o zmianie ustawy o podatku od towarów i usług oraz niektórych innych ustaw[9]. Ma na celu uszczelnienie systemu podatkowego i usprawnienie obrotu gospodarczego[10]. W wersji fakultatywnej KSeF wdrożono 1 stycznia 2022 r. Jego obligatoryjne wdrożenie stało się możliwe po otrzymaniu decyzji wykonawczej Rady (UE) 2022/1003 z 17.6.2022 r. upoważniającej Polskę do wprowadzenia szczególnego środka stanowiącego odstępstwo od dyrektywy 2006/112/WE. Stosowanie fakturowania elektronicznego w odniesieniu do wszystkich transakcji dokonywanych przez podatników, które wymagają wystawienia faktury (zwanego „szczególnym środkiem"), zostało ograniczone w czasie i może mieć miejsce od 1 stycznia 2024 r. do 31 grudnia 2026 r.

W Dzienniku Urzędowym Unii Europejskiej 25 marca 2025 r. opublikowano pakiet VAT in Digital Age (ViDA), jako dyrektywę Rady (UE) 2025/516[11]. Przepisy zwalniają kraje członkowskie od 2025 r. z obowiązku występowania do Komisji Europejskiej o zgodę na e-fakturowanie. Dodatkowo dyrektywa wprowadza od 1 lipca 2030 r. obowiązek wystawiania faktur transgranicznych w unijnej normie EN 16931. Od 1 stycznia 2035 r. standard UE e-fakturowania obejmie transakcje krajowe. Polska będzie miała czas do 2035 r. na dostosowanie swoich rozwiązań do standardów unijnych, z uwagi na okres przejściowy dla państw członkowskich, które

[8] K. Racka: *Big data – znaczenie, zastosowania i rozwiązania technologiczne*, „Zeszyty Naukowe PWSZ w Płocku, Nauki Ekonomiczne", t. XXIII, 2016.
[9] Dz.U. poz. 2076.
[10] <https://ksef.podatki.gov.pl/co-to-jest-ksef> (dostęp: 31.10.2025).
[11] Dyrektywa Rady (UE) 2025/516 z 11.3.2025, zmieniająca dyrektywę 2006/112/WE w odniesieniu do przepisów dotyczących VAT w epoce cyfrowej.

uzyskały zgodę na wprowadzenie e-fakturowania przed 1 stycznia 2024 r. Należy oczekiwać, że do 2035 r. będą prowadzone u nas prace zmierzające do umożliwienia konwersji struktury pliku XML[12] faktury ustrukturyzowanej FA(3) do normy EN 16931, ponieważ nie są to tożsame formaty. EN 16931 jest europejskim modelem semantycznym faktury elektronicznej, natomiast FA(3) polską strukturą XML przeznaczoną do obsługi faktur w KSeF. FA(3) nie jest implementacją EN 16931 – oba formaty funkcjonują równolegle i mają różne zastosowania.

Pierwotny termin wdrożenia obligatoryjnego KSeF był wyznaczony na 1 lipca 2024 r. W związku z koniecznością przeprowadzenia audytu i usunięcia problemów technicznych Ministerstwo Finansów zdecydowało o przesunięciu tego terminu. Ustawa z 9 maja 2024 r. zmieniająca ustawę o zmianie ustawy o podatku od towarów i usług przesunęła obowiązek wystawiania e-faktur na 1 lutego 2026 r.[13]

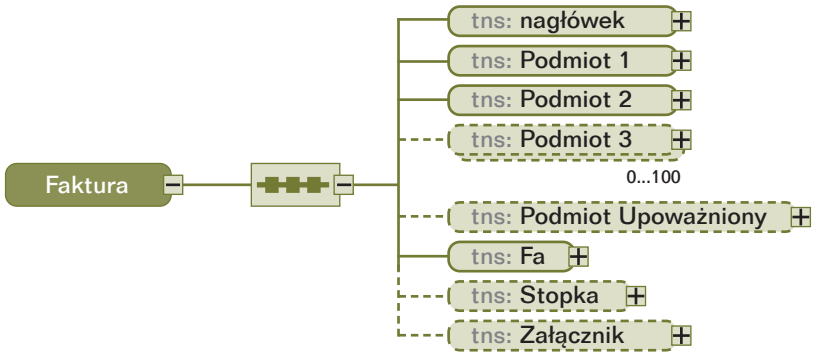
Definicja legalna faktury ustrukturyzowanej znajduje się w art. 2 pkt 32a ustawy z 11 marca 2004 r. o podatku od towarów i usług (dalej: ustawa o VAT). Zgodnie z tym przepisem, jest to faktura wystawiona przy użyciu Krajowego Systemu e-Faktur wraz z przydzielonym w tym systemie numerem identyfikacyjnym. Faktura ustrukturyzowana jest wystawiana i otrzymywana w ramach KSeF za pomocą oprogramowania interfejsowego, w postaci elektronicznej i zgodnie ze wzorem dokumentu elektronicznego w rozumieniu ustawy z 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne[14].

W myśl art. 106gb ust. 8 ustawy o VAT w brzmieniu obowiązującym od 1 lipca 2024 r. minister właściwy do spraw finansów publicznych udostępnia na elektronicznej platformie usług administracji publicznej wzór faktury ustrukturyzowanej. Struktura logiczna e-Faktury FA(3) została opublikowana 25 czerwca 2025 r. w Centralnym Repozytorium Wzorów Dokumentów Elektronicznych na ePUAP. Od 1 lutego 2026 r. zastąpiła strukturę logiczną e-Faktury FA(2).

Formatem pliku faktury ustrukturyzowanej jest XML, w którym pola (elementy) przyjmują charakter: obligatoryjny, opcjonalny lub fakultatywny. W przypadku pól obligatoryjnych zapisów dokonuje się obowiązkowo (np. NIP Podmiotu 1 – Sprzedawcy). Obligatoryjność pól co do zasady wynika z art. 106e ustawy o VAT, który określa co musi zawierać wystawiona faktura, niezależnie od tego, czy jest to dokument papierowy/elektroniczny, czy ustrukturyzowany. Pola opcjonalne to pola, które stają się obowiązkowe tylko wtedy, gdy zaistnieje dla nich określony warunek prawny lub transakcyjny. Jeśli ten warunek nie zachodzi, ich wypełnienie nie jest wymagane dla poprawności semantycznej pliku. Jako przykład można podać Pole

[12] XML (ang. Extensible Markup Language – rozszerzalny język znaczników) – uniwersalny język znaczników, przeznaczony do reprezentowania różnych danych w strukturalizowany sposób.
[13] Dz.U. poz. 852.
[14] tj. Dz.U. z 2024 r. poz. 1557, ze zm.

Rysunek 1. **Struktura logiczna faktury ustrukturyzowanej FA(3)**



Źródło: Wzór faktury ustrukturyzowanej opublikowany 25.6.2025 w Centralnym Repozytorium Wzorów Dokumentów Elektronicznych na ePUAP, struktura logiczna FA(3).

P_18A (Mechanizm Podzielonej Płatności – MPP). Pola fakultatywne to informacje nieobowiązkowe, które można dodać w celu ułatwienia obrotu gospodarczego. Należą do nich np. numery zamówień, dokumentów WZ, dane kontaktowe czy uwagi dla klienta.

Pola (elementy) opisane w prostokątach zakreślonych linią nieprzerywaną są obligatoryjne:
1. Nagłówek zawiera metadane systemowe i techniczne (np. kod, wariant, data wytworzenia pliku XML);
2. Podmiot 1 zawiera informacje, które charakteryzują podatnika sprzedawcę towaru lub usługi (np. NIP, pełna nazwa firmy lub imię i nazwisko, adres siedziby lub zamieszkania);
3. Podmiot 2 dotyczy danych, które charakteryzują nabywcę towaru lub usługi (np. NIP – tylko w wypadku nabywcy polskiego);
4. Fa zawiera dane dotyczące transakcji (np. nazwy towarów lub usług, ich ceny, ilości oraz stawki i kwoty podatku).

Pozostałe elementy opisane w prostokątach zakreślonych linią przerywaną są opcjonalne lub fakultatywne:
1. Podmiot 3 odnosi się do fakultatywnego elementu faktury (FA(3)), który pozwala na dodanie informacji o podmiotach trzecich, innych niż sprzedawca (Podmiot 1) i nabywca (Podmiot 2), a związanych z transakcją. Obejmuje to m.in. faktoring, grupy VAT, dodatkowych nabywców czy płatników;
2. Podmiot Upoważniony zawiera informacje, które charakteryzują podmiot upoważniony związany z fakturą, np. komornika. Jest to element opcjonalny;
3. Stopka obejmuje dane, np. nr KRS, wartość kapitału zakładowego, REGON. Jest to element fakultatywny faktury;
4. Załącznik stanowi ustrukturyzowaną część faktury w formacie XML, a nie osobny plik (np. PDF, JPG). Służy do przekazywania dodatkowych, specyficznych danych, takich jak szczegółowe zestawienia ilości, cen czy numerów punktów poboru energii, w formie tabelarycznej,

maksymalnie do 20 kolumn. Jest to fakultatywny element, który można dodać do faktury, jeśli przedsiębiorca zgłosi taką potrzebę w systemie e-Urząd Skarbowy.

Wdrażanie KSeF spotkało się z licznymi głosami krytycznymi, zgłaszanymi przez m.in. Rzecznika Małych i Średnich Przedsiębiorców[15], parlamentarzystów[16], Stowarzyszenie Księgowych w Polsce[17], przedstawicieli nauki[18]. Krytyka odnosiła się przede wszystkim do skali i kosztów obowiązków związanych z wdrożeniem KSeF, stopnia przygotowania przedsiębiorców i administracji do jego funkcjonowania, ryzyka wystąpienia problemów technicznych, a także bezpieczeństwa danych. Wskazywano również na potrzebę oceny, czy rozbudowany zakres informacji gromadzonych w KSeF rzeczywiście przekłada się na proporcjonalne zwiększenie efektywności zarządzania ryzykiem podatkowym i uszczelnianie systemu podatkowego. Wątpliwości dotyczące wdrożenia systemu nie ograniczały się zatem do kwestii technicznych. Obejmowały również zasadność nakładania na przedsiębiorców kolejnego obowiązku o charakterze raportowym, przygotowanie infrastruktury oraz zapewnienie odpowiednich mechanizmów na wypadek awarii systemu.

Od kilkunastu lat jest w Polsce widoczna zmiana podejścia do przekazywania informacji organom podatkowym, od raportowania w formie standardowej do raportowania w czasie rzeczywistym (określanego jako *real-time reporting*)[19]. To właśnie ta ostatnia cecha powoduje, że dane z faktury ustrukturyzowanej mogą być w sposób efektywny i skuteczny wykorzystane w ramach uczenia maszynowego. Wprowadzony od 1 stycznia 2018 r. Jednolity Plik Kontrolny (JPK-VAT) nie daje takich możliwości, ponieważ obowiązek jego sporządzania i przesyłania powstaje raz w miesiącu lub kwartale za wcześniejszy okres rozliczeniowy i nie zawiera tak szczegółowych danych, co faktura ustrukturyzowana, np. nazwy (rodzaju) towaru lub usługi[20].

Uczenie maszynowe i FA(3)

Struktura faktury ustrukturyzowanej FA(3), obowiązująca w KSeF od 1 lutego 2026 r., składa się z około 400 pól,

---

[15] <https://rzecznikmsp.gov.pl/mikroprzedsiebiorcy-potrzebuja-wiecej-czasu-na-wdrozenie-ksef/> (dostęp: 3.9.2026).

[16] Interpelacja nr 13534 w sprawie wstrzymania obligatoryjnego wdrożenia Krajowego Systemu e-Faktur (KSeF), <https://www.sejm.gov.pl/sejm10.nsf/interpelacja.xsp?typ=INT&nr=13534> (dostęp: 3.9.2026).

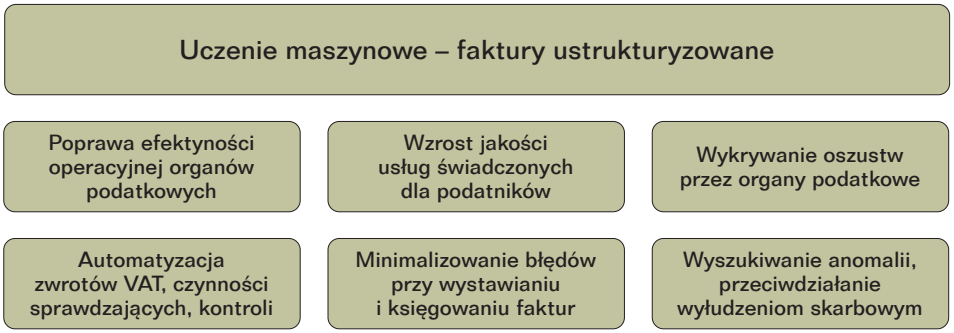
[17] <https://skwp.pl/aktualnosc/komunikat-ministerstwa-finansow-i-krajowej-administracji-skarbowej-w-sprawie-ksef-stanowisko-skwp/> (dostęp: 3.9.2026).

[18] P. Kobylski: *Krajowy System e-Faktur – bariery i wyzwania transformacji fakturowania*, „Palestra" nr 7/2025, s. 90-101; D. Nalazek, R. Mularczyk, R. Berliński: *Krajowy System e-Faktur – kolejny uciążliwy obowiązek raportowy czy milowy krok w kierunku cyfryzacji procesów zarządzania ryzykiem podatkowym*? „Przegląd Podatkowy" nr 5/2023, s. 36-47.

[19] Ch. Giles: *Expect the taxman of the future to take your money in real time. Special Report How we will live in 2050*, „Financial Times", 5.11.2019. Podobnie: M. Rydzewska, I. Majchrzak: *Cyfryzacja podatków – korzyści i wyzwania dla małych przedsiębiorstw i księgowych* [w:] A. Siedlecka, D. Guzal-Dec, J. Żurakowska-Sawa (red.), *Rozwój e-gospodarki wobec wyzwań pandemii COVID-19*, Biała Podlaska 2023.

[20] L. Syguła: *Krajowy System e-Faktur. Faktura ustrukturyzowana FA(3) jako narzędzie audytu podatkowego*, „Monitor Podatkowy" nr 1/2026, s. 37-41.

Rysunek 2. **Wykorzystanie uczenia maszynowego w analizie danych z faktury ustrukturyzowanej**



Źródło: Opracowanie własne.

z których 37 jest obowiązkowych (11 więcej niż w FA(2)). Nowy schemat zawiera dodatkowe pola, w tym sekcje Podmiot 3, a także rozbudowane dane dotyczące Grup Towarów i Usług (GTU), Polskiej Klasyfikacji Wyrobów i Usług (PKWiU), procedur i walut. Faktury są przesyłane przez ok. 2 mln czynnych podatników VAT oraz 755 tys. podmiotów zwolnionych[21]. Taki wolumen danych sprawia, że efektywne ich wykorzystanie będzie możliwe m.in. przy wykorzystaniu metod i narzędzi sztucznej inteligencji, ze szczególnym uwzględnieniem uczenia maszynowego. Działania te powinny być prowadzone z uwzględnieniem:

- poprawy efektywności operacyjnej organów podatkowych,
- wzrostu jakości usług świadczonych podatnikom,
- wykrywania oszustw przez organy podatkowe.

Poprawa efektywności organów podatkowych

Metody sztucznej inteligencji mogą zwiększać wewnętrzną wydajność operacyjną administracji podatkowej, zmniejszając zależność od czynności wykonywanych ręcznie i zwiększając ich dokładność. KSeF umożliwia pozytywne zmiany w obszarze podatkowym. W szczególności warto zwrócić uwagę na skrócenie podstawowego terminu zwrotu VAT z 60 na 40 dni[22]. Dostęp do faktur podatnika pozwala administracji usprawnić i przyspieszyć proces weryfikacji zasadności zwrotów VAT. Faktury ustrukturyzowane są przekazywane i rejestrowane w czasie rzeczywistym. Oznacza to, że w modelu online (podstawowym),

[21] Projekt ustawy o zmianie ustawy o podatku od towarów i usług oraz niektórych innych ustaw (Numer projektu: UD468).

[22] Art. 87 ust. 2 ustawy z 11.3.2004 o podatku od towarów i usług zmieniony ustawą z 16.6.2023 (Dz.U. z 2023 r. poz. 1598, ze zm., Dz.U. z 2024 r. poz. 852), która weszła w życie 1.2.2026.

po wystawieniu faktury przez sprzedawcę, trafia ona do systemu KSeF niemal natychmiast, stając się dostępną dla nabywcy. Taki obieg informacji pozwala na automatyczną weryfikację danych ze sprzedaży. Szacuje się, że obligatoryjny system KSeF będzie obsługiwał ponad 2,5 mld ustrukturyzowanych faktur rocznie[23]. Ten wolumen danych sprawia, że uczenie maszynowe może być jednym z rozwiązań wpływających na skuteczność modeli.

Podatek VAT podlega odliczeniu, gdy nabywane towary i usługi wykazują bezpośredni lub pośredni związek z działalnością opodatkowaną. Kluczowe jest, aby zakupione towary/usługi służyły sprzedaży opodatkowanej, co daje pełne prawo do obniżenia VAT należnego o naliczony. Przy braku związku ze sprzedażą opodatkowaną (np. zakupy na cele prywatne lub zwolnione) odliczenie nie przysługuje[24]. Uczenie maszynowe może pomóc w interpretacji złożonych opisów na fakturach; jest to przydatne przy weryfikacji, czy zakupy kwalifikują się do odliczeń.

Informacje z faktur ustrukturyzowanych (ewidencji zakupu i sprzedaży) pozwalają na analizę danych gospodarczych (cech – x) takich jak: sprzedaż, zakupy, podatek naliczony, podatek należny, wewnątrzwspólnotowa dostawa towarów, nabycie środków trwałych itd. Przykładowe cechy możliwe do zastosowania przy analizie zasadności zwrotu VAT przedstawia tabela 1 na s. 114.

Wyznaczenie zbioru uczącego to kluczowy etap w uczeniu maszynowym, polegający na przygotowaniu danych wejściowych oraz odpowiadających im pożądanych wartości wyjściowych, czyli etykiet/klas, na których model będzie się uczył. Jest to proces tworzenia reprezentatywnej próbki danych, zmierzający do poprawnego uogólnienia wiedzy i działania na nowych, nieznanych przykładach. Ewaluacja modelu pozwoli na określenie optymalnej liczby cech koniecznych do uzyskania jak najbardziej precyzyjnych wyników, aby osiągnąć jak największą precyzję w oznaczaniu zwrotów VAT, jako zasadnych lub niezasadnych.

Wzrost jakości świadczonych usług

Władze dążą do poprawy świadczenia usług publicznych zapewniając dostęp do sprawiedliwej i efektywnej administracji podatkowej. Sztuczna inteligencja pomaga usprawnić procesy, umożliwia szybszą komunikację, szybsze rozwiązania i bardziej przejrzyste interakcje z podatnikami.

Wdrożenie KSeF wymaga po stronie podatników nakładów inwestycyjnych w oprogramowanie finansowo-księgowe. Wystawianie faktur ustrukturyzowanych w formacie określonym przez Ministerstwo Finansów przyczyni się jednak do zapewnienia ich zgodności z wymogami wynikającymi z obowiązujących przepisów prawa, a także ograniczy ryzyko występowania błędów formalnych i nieprawidłowości w procesie fakturowania. KSeF jest systemem typu *clearance* (z ang. oczyszczenie, zwolnienie).

[23] Projekt ustawy o zmianie ustawy o podatku od towarów i usług oraz niektórych innych ustaw (nr projektu: UD468).
[24] Art. 86 ust. 1 ustawy z 11.3.2004 o podatku od towarów i usług (Dz.U. z 2025 r. poz. 775, t.j.).

Tabela 1. **Przykładowe cechy i etykiety zbioru uczącego się z faktury ustrukturyzowanej do zastosowania w uczeniu maszynowym**

Cecha	Zwrot 1	Zwrot 2	Zwrot 3
Wartość sprzedaży (x_1)	100	1000	200
Wartość zakupów (x_2)	1000	800	1000
Podatek naliczony (x_3)	230	184	230
Podatek należny (x_4)	23	230	46
WDT (x_5)	TAK	TAK	TAK
Inwestycje (x_6)	TAK	TAK	TAK
Zakup związany z PKD (x_7)	TAK	NIE	TAK
Faktura wykazana w JPK-VAT sprzedawcy (x_8)	NIE	TAK	TAK
Liczba pozycji faktury (x_9)	<10,100>	<10	>100
Częstotliwość współpracy z kontrahentem (liczba transakcji) (x_{10})	<5	<5	<5
Historia korekt (x_{11})	NIE	TAK	NIE
Sezonowość sprzedaży (x_{12})	TAK	TAK	TAK
Czas pomiędzy wystawieniem a zapłatą (x_{13})	10 dni	Brak zapłaty	30 dni
. ($x_{...}$)			
. ($x_{...}$)			
Etykieta (klasa) – zwrot	Zasadny	Niezasadny	Zasadny
Etykieta (klasa) – faktura	Prawidłowa	Fraudowa	Prawidłowa
Etykieta (klasa) – płatność	W terminie	Po terminie	Po terminie

Źródło: Opracowanie własne na podstawie struktury logicznej FA(3).

Faktury ustrukturyzowane są przesyłane do systemu, tam po walidacji fakturze jest nadawany numer KSeF, a odbiorca może ją pobrać z systemu. Model *clearance* powoduje, że KSeF jest „reaktywny" – użytkownik wprowadza dane, a system je waliduje. Dzięki uczeniu maszynowemu można przesunąć go w stronę systemu „proaktywnego" i wspierającego decyzje. Z perspektywy podatnika KSeF powinien „rozumieć", co on chce zrobić, zapobiegać błędom, automatyzować powtarzalne czynności. Uczenie maszynowe można wykorzystać w Aplikacji Podatnika KSeF, udostępnianej bezpłatnie przez Ministerstwo Finansów i systemach finansowo księgowych w celu generowania inteligentnych podpowiedzi przy wystawianiu faktur lub automatycznym uzupełnianiu danych kontrahenta. Algorytmy uczenia maszynowego pozwalają m.in. na:

- sugerowanie stawek VAT,
- wykrywanie nietypowych wartości (np. wysokiej kwoty na fakturze),
- sugerowanie nietypowej liczby pozycji faktury,
- wykrywanie niezgodności pomiędzy wartością netto i podatku,

• wykrywanie niestandardowych wzorców sprzedaży,
• wykrywanie duplikatów faktur,
• analizę sieci powiązań pomiędzy kontrahentami,
• ostrzeganie przed nietypowymi terminami płatności,
• wykrywanie nieprawidłowych numerów rachunków bankowych,
• niestosowanie mechanizmu podzielonej płatności.

Efektem wykorzystania uczenia maszynowego może być ograniczenie błędów. Jest to bardzo istotne zadanie, biorąc pod uwagę doświadczenia z wdrażania JPK_VAT, w ramach którego nie ma systemu „wczesnego ostrzegania", co prowadzi do występowania znacznej liczby i kwoty rozbieżności w wykazywanych przez podatników danych[25]. Kwota VAT zidentyfikowana z wykorzystaniem JPK_VAT jako „podejrzana" i przekazana do weryfikacji wynosiła w 2021 r. – 5,7 mld zł, w 2022 r. – 6,7 mld zł, w 2023 r. – 4,6 mld zł, a w 2024 r. 4,9 mld zł. Ministerstwo Finansów wskazuje, że przekazywane do weryfikacji organom podatkowym kwoty rozbieżności nie zawsze wynikają z nieprawidłowości związanych z niewłaściwym rozliczeniem podatku VAT. Część wartości wynika z pomyłek i błędów rachunkowych podatników, które nie mają wpływu na prawidłowość rozliczenia podatku, a korekty JPK_VAT składane po weryfikacji dokonanej przez urzędy skarbowe nie wpływają na kwoty podatku VAT.

Zastosowanie uczenia maszynowego pozwoli także na zautomatyzowanie procesu klasyfikacji i księgowania faktur, dzięki przypisywaniu ich do właściwych kategorii kosztowych oraz sugerowaniu kont księgowych.

Wykrywanie oszustw podatkowych

Krajowa Administracja Skarbowa (KAS) wdrożyła w 2019 r. projekt Silnik Analiz (SA), wykorzystujący innowacyjne metody analityczne (m.in. modelowanie statystyczno-ekonometryczne, uczenie maszynowe)[26]. Na podstawie jego środowiska KAS identyfikuje rozbieżności w rozliczeniu podatku od towarów i usług, które trafiają do urzędów skarbowych oraz urzędów celno-skarbowych. Mając na uwadze dotychczasowe doświadczenia, a także rozwiązania stosowane w innych krajach, zasadne jest prowadzenie dalszych prac w tym obszarze z wykorzystaniem danych z faktur ustrukturyzowanych.

Rozwiązania stosowane we Włoszech wydają się interesujące i możliwe do zastosowania w Polsce. Włochy są liderem we wdrażaniu obowiązkowego fakturowania elektronicznego i uczenia maszynowego w celu zwalczania oszustw fakturowych i unikania płacenia podatku VAT. Po wprowadzeniu obowiązkowego e-fakturowania B2B w 2019 roku włoska administracja skarbowa (Agenzia delle Entrate) wykorzystuje zautomatyzowane systemy do skanowania milionów

[25] Odpowiedź na Interpelację nr 10254 w sprawie raportowania JPK_VAT, pismo Ministerstwa Finansów, DAK1.054.1.2025, z 30.6.2025.
[26] <https://www.gov.pl/web/kas/analizy-rozliczen-vat-z-wykorzystaniem-jpkvat-stir-i-silnika-analiz>

ustrukturyzowanych faktur XML w celu wykrywania anomalii w czasie rzeczywistym. To duża zaleta tego rozwiązania[27]. Stosuje ona zaawansowane algorytmy do analizy danych z centralnego systemu wymiany faktur (Sistema di Interscambio – SDI). System analityczny VeRa (Vettori di Rischio) weryfikuje zaś e-faktury z danymi bankowymi, informacjami celnymi i ewidencją nieruchomości. VeRa pozwala na wykrywanie niezgodnych faktur lub niezadeklarowanego dochodu.

Efektywne wykorzystanie danych z faktury ustrukturyzowanej w procesie uczenia maszynowego w Polsce powinno pozwolić na stworzenie modeli sygnalizujących m.in. nagły wzrost wartości sprzedaży bez analogicznego wzrostu zakupów, bardzo wysokie zwroty VAT przy niewielkiej sprzedaży krajowej, liczne faktury wystawiane pomiędzy tą samą grupą podmiotów, faktury niewykazane w ewidencji sprzedaży, zakupy niezwiązane z deklarowanym profilem działalności, brak płatności za faktury, krótkotrwałe pojawienie się nowych kontrahentów generujących wysokie obroty. Skuteczność tych działań powinna być znacząca z uwagi na raportowanie danych w czasie rzeczywistym oraz stosowanie tzw. niezmiennych stempli czasowych, które pozwalają nadać każdej fakturze unikalny numer w momencie jej przesłania. To powoduje, że niemożliwe jest antydatowanie faktur i błędne rozpoznawanie dat powstawania obowiązków podatkowych w VAT i podatkach dochodowych.

Bardzo interesującym rozwiązaniem wdrożonym w Aplikacji Podatnika KSeF jest możliwość oznaczania faktur oszukańczych (tzw. scamowych). Zgłaszając fakturę w Aplikacji Podatnika KSeF należy podać jedną z pięciu przesłanek uznania jej przez podatnika za scamową: próba wyłudzenia nienależnej zapłaty, niechciane treści reklamowe, podejrzane linki, masowa wysyłka faktur, podszywanie się pod instytucję[28].

Lista faktur scamowych może posłużyć do wyeliminowania ich z rozliczeń wszystkich przedsiębiorców, pod warunkiem uznania, że rzeczywiście istnieją okoliczności wymienione w art. 88 ust. 3a ustawy o VAT.

Zakończenie

Podsumowując, integracja uczenia maszynowego z Krajowym Systemem e-Faktur (fakturą ustrukturyzowaną FA(3)) otwiera nowe perspektywy zarówno dla przedsiębiorców, jak i Krajowej Administracji Skarbowej. Wykorzystanie nowoczesnych technologii pozwala nie tylko usprawnić procesy wewnętrzne i zwiększyć efektywność operacyjną, lecz także znacząco poprawić komfort i bezpieczeństwo podatników w codziennym funkcjonowaniu. Jednocześnie zaawansowane algorytmy analityczne stwarzają realną szansę na skuteczniejsze wykrywanie nieprawidłowości

[27] W. A. Radhi, A. Hamdan, R. Binsaddig: *Assessing the role of artificial intelligence (AI) on tax fraud detection. In Business Development via AI and Digitalization: Volume 1*, Springer Nature Switzerland 2024, s. 359-364.

[28] *Podręcznik użytkownika Krajowy System e-Faktur – Aplikacja Podatnika*, wersja 2.6, Ministerstwo Finansów, s. 119, <https://ksef.podatki.gov.pl/media/qocjn3ai/podrecznik-uzytkownika-aplikacji-podatnika-ksef-20_06082026.pdf>.

i ograniczanie zjawisk, takich jak oszustwa podatkowe.

W nadchodzących latach kluczowe będzie jednak odpowiednie zbalansowanie innowacyjności z przejrzystością i zaufaniem użytkowników do systemu. Tylko spójne podejście, łączące rozwój technologiczny z dbałością o potrzeby interesariuszy, pozwoli w pełni wykorzystać potencjał, jaki niesie ze sobą cyfryzacja procesów fakturowania.

dr **LESZEK SYGUŁA**
doradca podatkowy,
Uniwersytet WSB Merito we Wrocławiu
ORCID: 0009-0001-6020-7332

Słowa kluczowe: uczenie maszynowe, faktura ustrukturyzowana, Krajowy System e-Faktur, analiza danych

Bibliografia:

1. Begum T.: *Predictive analytics for machine learning and deep learning. In Handbook of big data research methods*, Edward Elgar Publishing 2023.
2. Brzeziński B.: *Prawo podatkowe. Zagadnienia teorii i praktyki*, Toruń 2017.
3. Giles Ch.: *Expect the taxman of the future to take your money in real time. Special Report How we will live in 2050*, „Financial Times", 5.11.2019.
4. Goodfellow I., Bengio Y., Courville A.: *Deep Learning*. MIT Press, Cambridge 2016.
5. Kędzierski M. A.: *Propozycja wykorzystania uczenia przez wzmacnianie w celu optymalizowania podejmowania decyzji w zakresie przeciwdziałania praniu pieniędzy oraz finansowania terroryzmu (część 2)*, „Nowoczesne Systemy Zarządzania" nr 4/2023.
6. Kobylski P.: *Krajowy System e-Faktur – bariery i wyzwania transformacji fakturowania*, „Palestra" nr 7/2025.
7. Menon S.: *How can AI drive audits*, ISACA 2021.
8. Nalazek D., Mularczyk R., Berliński R.: *Krajowy System e-Faktur – kolejny uciążliwy obowiązek raportowy czy milowy krok w kierunku cyfryzacji procesów zarządzania ryzykiem podatkowym*? „Przegląd Podatkowy" nr 5/2023.
9. Paliński A.: *Metody uczenia maszynowego w prognozowaniu niewypłacalności*, „Studia Ekonomiczne" nr 358/2018.
10. Prędkiewicz K., Biegun K.: *AI – sztuczna inteligencja w finansach przedsiębiorstw* [w:] *Finanse przedsiębiorstw*, M. Sierpińska i P. Szczepanowski (red.), Polska Akademia Nauk 2024.
11. Racka K.: *Big data - znaczenie, zastosowania i rozwiązania technologiczne*, „Zeszyty Naukowe PWSZ w Płocku, Nauki Ekonomiczne", t. XXIII, 2016.
12. Radhi W. A., Hamdan A., Binsaddig R.: *Assessing the role of artificial intelligence (AI) on tax fraud detection. In Business Development via AI and Digitalization: Volume 1*, Cham: Springer Nature Switzerland 2024.

13. Rydzewska M., Majchrzak I.: *Cyfryzacja podatków – korzyści i wyzwania dla małych przedsiębiorstw i księgowych* [w:] A. Siedlecka, D. Guzal-Dec, J. Żurakowska-Sawa (red.), *Rozwój e-gospodarki wobec wyzwań pandemii COVID-19*, Biała Podlaska 2023.
14. Sikorski W.: *Sztuczna inteligencja, uczenie maszynowe i analiza big data w energetyce przyszłości*, „Kierunek Energetyka" nr 3/2024.
15. Syguła L.: *Krajowy System e-Faktur. Faktura ustrukturyzowana FA(3) jako narzędzie audytu podatkowego*, „Monitor Podatkowy" nr 1/2026.
16. Wróbel M., Wojda P.: *Możliwości optymalizacji procesu transportowego z wykorzystaniem sztucznej inteligencji*, „Journal of TransLogistics" nr 14/2018.

ABSTRACT

Potential Applications of Machine Learning in the National e-Invoicing System (KSeF) – Data Analysis from Structured Invoice

The experience with knowledge-based systems relying on predefined, hard-coded rules, has led to expectations that artificial intelligence agents should be capable of acquiring knowledge autonomously by extracting patterns from raw data. This capability is known as machine learning. Machine learning is the study of algorithms and systems that improve their knowledge and performance through experience. It is a subfield of artificial intelligence that seeks to emulate intelligent behaviour by identifying relationships in training data, for example, by detecting anomalies. Finance is one of the sectors that adopts artificial intelligence tools most rapidly. Solutions such as machine learning, including supervised, unsupervised and reinforcement learning, help address challenges related to analysing large datasets and extraction of meaningful insights. The National e-Invoicing System (Polish: *Krajowy System e-Faktur*, KSeF), as a dataset held by the Head of the National Revenue Administration, constitutes a big data repository. Accordingly, it will be subject to typical big data analyses that involve data mining, business analytics, diagnostics and reporting. Application of machine learning methods, in turn, will allow for using the KSeF dataset to identify patterns, generate predictions and support decision-making.

Leszek Syguła, PhD, Tax Adviser, WSB Merito University in Wrocław,
ORCID: 0009-0001-6020-7332

Key words: machine learning, structured invoice, the National e-Invoicing System (KSeF), data analysis