

ELEKTRONICZNA PLATFORMA WYCENY NIERUCHOMOŚCI

SYGN. WIB PAB 14/2020



Raport opracowany na zlecenie Programu Analityczno-Badawczego
Fundacji Warszawski Instytut Bankowości

Warszawa, lipiec 2020

WIB

PROGRAM
ANALITYCZNO
BADAWCZY

O raporcie

Raport „**Elektroniczna platforma wyceny nieruchomości**” został opracowany na zlecenie Programu Analityczno-Badawczego Fundacji Warszawski Instytut Bankowości

Autorzy

Raport przygotowany przez zespół w składzie:

Prof. dr hab. Krzysztof Jajuga

– profesor nauk ekonomicznych, kierownik Katedry Inwestycji Finansowych i Zarządzania Ryzykiem (KIFiZR), a także dyrektor Instytutu Zarządzania Finansami oraz wykładowca akademicki Uniwersytetu Ekonomicznego we Wrocławiu, Politechniki Warszawskiej, jak również uczelni za granicą (m.in. Jiao Tong University w Szanghaju). Doktor honorowy Uniwersytetu Ekonomicznego w Krakowie i profesor honorowy Politechniki Warszawskiej. W swojej pracy naukowo-badawczej koncentruje się wokół finansów, zarządzania ryzykiem oraz statystyki i ekonometrii. Autor 450 publikacji. Wypromował 40 doktorów.

Jest przewodniczącym Komitetu Statystyki i Ekonometrii Polskiej Akademii Nauk, zastępcą przewodniczącego Komitetu Nauk Ekonomicznych Polskiej Akademii Nauk, członkiem prezydium Komitetu Nauk o Finansach Polskiej Akademii Nauk, członkiem Centralnej Komisji do Spraw Stopni i Tytułów oraz Rady Doskonałości Naukowej. Ponadto jest twórcą i prezesem CFA Society Poland. Współpracuje z licznymi instytucjami finansowymi i przedsiębiorstwami.

Był członkiem Rady Naukowej Narodowego Banku Polskiego i Rady Giełdy Papierów Wartościowych w Warszawie. Pełnił i pełni obowiązki w licznych radach nadzorczych (w sumie kilkanaście).

Prof. dr hab. Andrzej Bąk

– profesor nauk ekonomicznych w Katedrze Ekonometrii i Informatyki Wydziału Ekonomii i Finansów Uniwersytetu Ekonomicznego we Wrocławiu. Absolwent Akademii Ekonomicznej we Wrocławiu (1986).

Prodziekan ds. nauki i przewodniczący Wydziałowej Komisji ds. Nauki (2005–2008 i 2008–2012) na Wydziale Ekonomii, Zarządzania i Turystyki w Jeleniej Górze. Członek Senatu Uniwersytetu Ekonomicznego we Wrocławiu w kadencjach 2012–2016 i 2016–2020. Członek Komitetu Redakcyjnego czasopisma „Informatyka Ekonomiczna” (od 2011), redaktor naczelny i członek Rady Naukowej czasopisma „Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu” (od 2012), członek Kolegium Redakcyjnego czasopisma „Argumenta Oeconomica” (od 2013).

Autor i współautor wielu monografii i artykułów naukowych poświęconych zastosowaniom metod statystycznej analizy wielowymiarowej w badaniach zjawisk ekonomicznych oraz pakietów obliczeniowych dla programu CRAN R (conjoint, DiscreteChoice, MaxDiff, dcMNM, pllord). Kierował projektem badawczym MNiSW: *Pomiar, analiza i wizualizacja preferencji ujawnionych i wyrażonych z wykorzystaniem metod wielowymiarowej analizy statystycznej i programu R* (2009–2012).

Członek stowarzyszeń naukowych: Sekcji Klasyfikacji i Analizy Danych PTS, Komisji Nauk Ekonomicznych Oddziału PAN we Wrocławiu.

Dr Anita Makowska

– doktor nauk ekonomicznych w dyscyplinie finanse (Uniwersytet Ekonomiczny we Wrocławiu). Absolwentka Uniwersytetu Ekonomicznego we Wrocławiu na kierunku finanse i bankowość, specjalność nieruchomości z tytułem magistra oraz Politechniki Wrocławskiej na kierunku zarządzanie i marketing, specjalność zarządzanie przedsiębiorstwem z tytułem magistra inżyniera. Sekretarz Komitetu Statystyki i Ekonometrii Polskiej Akademii Nauk. Redaktor czasopisma naukowego „Argumenta Oeconomica”. Członek komitetu redakcyjnego czasopisma „Świat Nieruchomości” (World of Real Estate Journal).

Autorka wielu publikacji naukowych z dziedziny nieruchomości i rynku nieruchomości. Współpracowała z licznymi firmami, prowadząc szkolenia i przygotowując analizy rynku. Od dziesięciu lat wykładowca akademicki na Uniwersytecie Ekonomicznym we Wrocławiu takich przedmiotów, jak m.in. inwestowanie i instrumenty na rynku nieruchomości, finanse, risk management in financial markets, real estate market, zaawansowane zagadnienia wyceny, usługi finansowe dla przedsiębiorstw, financial mathematics.

Od 2014 roku zawodowo zajmuje się analizą rynku nieruchomości. Szczególnym przedmiotem jej badań jest wrocławski rynek nieruchomości mieszkaniowych.

O programie

Program Analityczno-Badawczy przy Fundacji WIB powstał w 2019 roku jako odpowiedź na potrzeby sektora bankowego w zakresie analizy zjawisk, tworzenia opracowań i porządkowania wiedzy w obszarach cyberbezpieczeństwa i nowych technologii, a także szeroko rozumianego otoczenia sektora bankowego, kształtującego warunki działania banków w Polsce. Prace analityczno-badawcze w ramach programu realizowane są pod kątem możliwości praktycznego wykorzystania ich wyników w celu rozwoju sektora bankowego, podnoszenia poziomu bezpieczeństwa usług oraz kreowania wartości dla klientów bankowości. Pab wib kładzie duży nacisk na rozwój współpracy ze środowiskami akademickimi i eksperckimi, poszukując synergii w zakresie zainteresowań badawczych autorów oraz potrzeb rozwojowych sektora bankowego.

W ramach programu realizowane są analizy i badania w następujących obszarach:

-  **Nowe technologie i cyberbezpieczeństwo**
-  **Zdolność banków do finansowania gospodarki**
-  **Bankowość spółdzielcza**
-  **Rynek nieruchomości**
-  **Zielony ład i finansowanie energetyki odnawialnej**
-  **Finansowanie projektów innowacyjnych**

Więcej na temat działalności programu na stronie www.pab.wib.org.pl

Kontakt

dr Andrzej Banasiak
Koordynator Programu
m: 696 405 104
e: abanasiak@wib.org.pl

Jacek Gieorgica
m: 603 626 254
e: jgieorgica@wib.org.pl

Warszawski Instytut Bankowości
00-394 Warszawa, ul. Solec 38, lok 104
t: (22) 182 31 70
e: pab@wib.org.pl

Agnieszka Nierodka
m: 607 484 391
e: anierodka@wib.org.pl

dr Tomasz Pawlonka
m: 505 917 778
e: tpawlonka@wib.org.pl

Spis treści

	Streszczenie kierownicze	5
	Wprowadzenie – cel i interesariusze projektu	8
	ROZDZIAŁ I. Wycena nieruchomości – metody i systemy	10
	1.1 Regulacje prawne	11
	1.2 Metody analizy statystycznej rynku – ujęcie klasyczne	11
	1.3 Metody automatycznej wyceny nieruchomości	12
	1.4 Systemy automatycznej wyceny nieruchomości	14
	ROZDZIAŁ II. Dostępne bazy danych	15
	2.1 Identyfikacja dostępnych baz danych	16
	2.2 Określenie zmiennych wejściowych i weryfikacja ich dostępności w bazach danych	17
	2.3 Metody wyodrębniania zbiorów nieruchomości podobnych	20
	ROZDZIAŁ III. Zbiór potencjalnych metod wyceny	22
	3.1 Opis zbioru potencjalnych metod wyceny	23
	3.2 Określenie jakości funkcjonowania metody wyceny	24
	3.3 Badania w zakresie zastosowania metod analizy statystycznej rynku i możliwości ich implementacji w programie CRAN R – schemat i wyniki badań	25
	3.4 Struktura projektowanego systemu automatycznej wyceny nieruchomości	30
	3.5 Funkcjonalność systemu dla różnych interesariuszy	33
	3.6 Spełnianie wymogów regulacyjnych przez system	35
	ROZDZIAŁ IV. Dalsze proponowane działania	39
	Literatura	41



Streszczenie kierownicze





Prezentowany raport dotyczy koncepcji potencjalnego systemu automatycznej wyceny nieruchomości.

Celem projektu jest opracowanie szczegółowej koncepcji modelu dla systemu automatycznej wyceny nieruchomości mieszkaniowych z zastosowaniem zaawansowanych metod statystycznych i wstępny test niektórych elementów systemu.

Informacje o cenach (wartościach) nieruchomości są istotne dla banków, które udzielają kredytów mieszkaniowych zabezpieczonych hipotecznie. Użyteczność dla banków, jeśli chodzi o koncepcję systemu zawartą i częściowo testowaną w projekcie, wynika z następujących informacji, z których bank może korzystać:

- ▶ wstępna wycena nieruchomości w momencie wydawania formularza informacyjnego;
- ▶ wycena nieruchomości w momencie podpisywania umowy kredytu (weryfikacja dostarczonego operatu szacunkowego);
- ▶ weryfikacja wartości nieruchomości określanej przez rzeczoznawcę w dowolnym momencie;
- ▶ wycena nieruchomości jako przedmiotu zabezpieczenia w sytuacji wystąpienia opóźnień w obsłudze kredytu;
- ▶ ocena poziomu zabezpieczenia ekspozycji kredytowej;
- ▶ wartość portfela zabezpieczeń kredytów w dowolnym momencie wyceny, jak również automatyczna informacja o tej wartości w ustalonych odstępach czasowych, np. raz w miesiącu;
- ▶ wycena *cover pool* w sytuacji transferu wierzytelności z banku uniwersalnego do banku hipotecznego (emitenta listów zastawnych);
- ▶ ocena poziomu zabezpieczenia ekspozycji kredytowej portfela.

Po rozszerzeniu systemu o podmoduł analityczno-prognostyczny bank będzie mógł korzystać z następujących dodatkowych elementów jego funkcjonalności:

- ▶ wyznaczenie prognozy wartości pojedynczej nieruchomości;
- ▶ wyznaczenie przyszłej wartości pojedynczej nieruchomości przy założeniu scenariusza kształtowania się sytuacji rynkowej – wstępnie prze-

widuje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych (tzw. *stress*);

- ▶ wyznaczenie prognozy wartości portfela zabezpieczeń;
- ▶ wyznaczenie przyszłej wartości portfela zabezpieczeń przy założeniu scenariusza kształtowania się sytuacji rynkowej – wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych (tzw. *stress*).

Koncepcja przedstawiona w projekcie dotyczy systemu automatycznej wyceny nieruchomości mieszkaniowych. Spełnione są wymogi formalne, gdyż wycena sporządzana jest z zastosowaniem metody analizy statystycznej rynku, zaliczanej do podejścia porównawczego. Zastosowane są w niej metody uczenia maszynowego, które pozwalają na wycenę z wykorzystaniem bazy danych o nieruchomościach. Porównanie różnych dostępnych baz danych doprowadziło do wybrania bazy AMRON, jako tej, która spełnia warunki jakości danych.

W raporcie przedstawione są kluczowe elementy projektowanego systemu automatycznej wyceny nieruchomości mieszkaniowych. Są nimi:

- ▶ zmienne – charakterystyki nieruchomości – przyjmowane przy wycenie; raport przedstawia tę listę;
- ▶ metody wyodrębniania klas nieruchomości podobnych; w raporcie do badań wybrana została metoda k-średnich, ale w projektowanym systemie wybór będzie dokonany z uwzględnieniem różnych kryteriów;
- ▶ metody wyceny z zestawu metod uczenia maszynowego; w raporcie do badań wybrane zostało sześć metod, ale w projektowanym systemie wybór będzie z większej ich grupy;
- ▶ wskaźniki jakości metod wyceny.

Badania testowe przeprowadzone w projekcie dotyczyły tych sześciu różnych metod uczenia maszynowego, które potencjalnie mogą być wykorzystane w automatycznej wycenie nieruchomości mieszkaniowych. Były one wykonane w odniesieniu do sześciu różnych miast na podstawie bazy danych o nieruchomościach AMRON, z uwzględnieniem (w odniesieniu do niektórych zbiorów) również podziału na zbiory nieruchomości podobnych.



Istotna część raportu zawiera propozycję struktury projektowanego systemu automatycznej wyceny nieruchomości. Obejmuje on trzy moduły: moduł wejściowy, moduł operacyjny (w ramach którego są: podmoduł wyceny, podmoduł analityczno-prognostyczny, podmoduł archiwizujący) oraz moduł wyjściowy. Oprócz standardowej wyceny, system może wykonywać operacje zawarte w module analityczno-prognostycznym, takie jak: wyznaczenie dla dowolnej wycenianej nieruchomości mieszkaniowej przedziału wyceny, wyznaczenie prognozy wartości nieruchomości, wyznaczenie prognozy przeciętnej wartości nieruchomości na dowolnym lokalnym rynku, wyznaczenie parametru zmienności cen dowolnej nieruchomości, wyznaczenie

prognozy wartości zabezpieczenia dla portfela kredytów, analiza *ex post* zmian wartości pojedynczej nieruchomości na podstawie danych historycznych zarchiwizowanych w systemie, analiza *ex post* zmian wartości portfela zabezpieczeń na podstawie danych historycznych zarchiwizowanych w systemie.

Raport zawiera opis funkcjonalności tego systemu dla różnych grup interesariuszy, przede wszystkim dla banków, jak również spełnianie warunków regulacyjnych przez system.



Wprowadzenie – cel i interesariusze projektu





Celem projektu jest opracowanie szczegółowej koncepcji modelu dla systemu automatycznej wyceny nieruchomości mieszkaniowych z zastosowaniem zaawansowanych metod statystycznych i wstępny test niektórych elementów systemu.

Nie ulega wątpliwości, że wycena nieruchomości mieszkaniowych jest jedną z podstawowych aktywności związanych z rynkiem nieruchomości. Informacje o cenach (wartościach) nieruchomości są istotne dla wielu interesariuszy. Przede wszystkim są nimi banki, udzielające kredytów mieszkaniowych zabezpieczonych hipotecznie. Inni interesariusze, którzy będą potencjalnymi użytkownikami systemu wyceny nieruchomości to:

- ▶ gospodarstwa domowe, zainteresowane wartością nabywanej lub posiadanej nieruchomości;
- ▶ inwestorzy indywidualni i instytucjonalni na rynku nieruchomości;
- ▶ decydenci gospodarczy (na poziomie centralnym i lokalnym), prowadzący politykę gospodarczą, w szczególności politykę mieszkaniową;
- ▶ decydenci gospodarczy, analizujący koniunkturę gospodarczą;
- ▶ deweloperzy, podejmujący decyzje biznesowe dotyczące budowy mieszkań;
- ▶ instytucje badawcze, opracowujące raporty o rynku nieruchomości;
- ▶ instytucje naukowe, prowadzące badania naukowe dotyczące nieruchomości.



Wycena nieruchomości – metody i systemy





1.1 Regulacje prawne

Polskie regulacje dotyczące rynku nieruchomości, w tym wyceny nieruchomości, są określone na trzech poziomach:

- ▶ poziom I: Ustawa z dnia 21 sierpnia 1997 r. o gospodarce nieruchomościami (Dz.U. z 2018 r., poz. 2204 ze zm.);
- ▶ poziom II: Rozporządzenie Rady Ministrów z dnia 21 września 2004 r. w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego (Dz.U. Nr 207, poz. 2109 ze zm.);
- ▶ poziom III: Powszechne Krajowe Zasady Wyceny i Noty interpretacyjne.

Art. 152 pkt. 2 Ustawy o gospodarce nieruchomościami określa:

„Wyceny nieruchomości dokonuje się przy zastosowaniu podejść: porównawczego, dochodowego lub kosztowego, albo mieszanego, zawierającego elementy podejść poprzednich”.

Przy tym wycena nieruchomości mieszkaniowych zazwyczaj dokonywana jest za pomocą podejścia porównawczego.

Art. 153 pkt. 1 Ustawy o gospodarce nieruchomościami określa:

„Podejście porównawcze polega na określeniu wartości nieruchomości przy założeniu, że wartość ta odpowiada cenom, jakie uzyskano za nieruchomości podobne, które były przedmiotem obrotu rynkowego. Ceny te koryguje się ze względu na cechy różniące nieruchomości podobne do nieruchomości wycenianej oraz uwzględnia się zmiany poziomu cen wskutek upływu czasu. Podejście porównawcze stosuje się, jeżeli są znane ceny i cechy nieruchomości podobnych do nieruchomości wycenianej”.

Z kolei art. 4 Ustawy o gospodarce nieruchomościami określa nieruchomość podobną jako „nieruchomość, która jest porównywalna z nieruchomością stanowiącą przedmiot wyceny, ze względu na położenie, stan prawny, przeznaczenie, sposób korzystania oraz inne cechy wpływające na jej wartość”.

Przy określaniu podobieństwa nieruchomości pod uwagę można brać różne cechy, takie jak np. lokalizacja, technologia, liczba pokoi itd.

Jeśli chodzi o metody stosowane w podejściu porównawczym, to Rozporządzenie Rady Ministrów w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego w paragrafie 4 pkt 2 określa:

„W podejściu porównawczym stosuje się metodę porównywania parami, metodę korygowania ceny średniej albo metodę analizy statystycznej rynku”.

W literaturze dotyczącej wyceny podkreśla się, biorąc pod uwagę kryteria naukowe, że metoda korygowania ceny średniej zawiera niedostatki metodyczne. Zdaniem pewnej grupy rzeczoznawców majątkowych, nie powinna być stosowana. Zdecydowanie najczęściej stosowana jest metoda porównywania parami. Rzadziej w procesie wyceny sięga się zaś po metody zaliczane do grupy metod analizy statystycznej rynku, które wymagają stosowania (nie zawsze skomplikowanych) metod statystycznych.

Rozporządzenie Rady Ministrów w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego w paragrafie 4 pkt 5 określa:

„Przy metodzie analizy statystycznej rynku przyjmuje się zbiór cen transakcyjnych właściwych do określania wartości nieruchomości reprezentatywnych [...]. Wartość nieruchomości określa się przy użyciu metod stosowanych do analiz statystycznych”.

Takie sformułowanie w regulacjach daje szerokie możliwości co do stosowania różnych metod statystycznych.

1.2 Metody analizy statystycznej rynku – ujęcie klasyczne

Metody analizy statystycznej rynku znane są również pod nazwą metody hedoniczne (*hedonic methods*) lub metody regresji hedonicznej. Pierwsze próby zastosowania tych metod do analizy cen nieruchomości (gruntowych) miały miejsce prawie 100 lat temu [por. Haas (1922)]. Idea metod hedonicznych polega na przyjęciu naturalnego założenia, że cena jakiegokolwiek dobra (w tym nieruchomości) jest zależna od cech (charakterystyk) tego dobra.

Metody statystyczne wyceny (metody analizy statystycznej rynku) sprowadzają się do wyznaczenia funkcji danej w następującej ogólnej postaci:

$$P = f(X_1, X_2, \dots, X_m, \epsilon) \quad 1$$

lub:

$$\ln P = f(X_1, X_2, \dots, X_m, \epsilon) \quad 2$$

Gdzie:

P – wartość nieruchomości;

$X_1, X_2, \dots, X_m$ – zmienne charakteryzujące nieruchomość;

f – funkcja zależności;

ϵ – składnik losowy.

Powyższy model hedoniczny wskazuje na funkcję uzależniającą wartość nieruchomości od zmiennych (cech) charakteryzujących tę nieruchomość, przy czym – z uwagi na składnik losowy – funkcja ta ma charakter aproksymacyjny.

Funkcja dana wzorem (1) lub (2) może mieć w zasadzie dowolną postać. Często – ze względu na prostotę i walory interpretacyjne – przyjmowana jest funkcja liniowa, historycznie postać ta była najczęściej stosowana w metodach hedonicznych.

Proces zastosowania metod hedonicznych do wyceny nieruchomości przebiega w dwóch etapach. Pierwszy polega na estymacji (oszacowaniu) funkcji danej wzorem (1) lub (2) na podstawie dostępnych danych zawartych w bazach danych o nieruchomościach i cenach transakcji na rynku nieruchomości (ewentualnie danych o wycenach nieruchomości dokonywanych przez rzeczoznawców). Drugi etap polega na wycenie rozpatrywanej – przez użytkownika metody – nieruchomości przez podstawienie jej charakterystyk jako wartości zmiennych $X_1, X_2, \dots, X_m$ do prawej strony wzoru (1) lub (2) i na tej podstawie wyznaczenie wartości nieruchomości (lewa strona wzoru).

Z punktu widzenia statystyki metody te są proste. Jednak skuteczność ich zastosowania zależy od kilku determinant. Są nimi przede wszystkim:

1. Dobór nieruchomości podobnych, dla których dane są wykorzystane do estymacji funkcji.
2. Wybór zmiennych objaśniających w funkcji.
3. Wybór postaci funkcji.
4. Wybór metody estymacji funkcji.

Spośród powyżej wymienionych czterech determinant dwie pierwsze mają charakter merytoryczny, dotyczą nieruchomości, zaś dwie ostatnie mają charakter metodyczny, dotyczą procesu analizy statystycznej. Czynniki pierwsze wskazują, że może (zazwyczaj) być wiele funkcji wyceny, z których każda przeznaczona jest do wyceny klasy nieruchomości podobnych. Jeśli chodzi o czynnik trzeci, to w standardowych metodach analizy statystycznej rynku funkcja zależności jest liniowa, mamy zatem do czynienia z klasycznym modelem regresji liniowej. W modelu tym liniowa funkcja regresji szacowana jest na podstawie danych historycznych dotyczących cen transakcyjnych nieruchomości oraz charakterystyk tych nieruchomości.

1.3 Metody automatycznej wyceny nieruchomości

Bardzo dynamiczny rozwój technologii informatycznej w ostatnim ćwierćwieczu, a zwłaszcza w ostatnich dziesięciu latach, spowodował wzrost zainteresowania metodami (a w szczególności algorytmami) wyznaczania funkcji zależności wartości nieruchomości od czynników wpływających na tę wartość. W szczególności duże zainteresowanie badaczy i praktyków budzą modele i metody, w których, po pierwsze, w miejsce funkcji liniowej (nie zawsze dobrze przybliżającej zależność między zmiennymi) zastosowanie ma skomplikowana funkcja nieliniowa, po drugie, stosowane są duże zbiory danych historycznych. Są to tzw. metody automatycznej wyceny nieruchomości (*Automated Valuation Methods*).

Jest pewna liczba określeń tych metod. Dla porządku przedstawimy najważniejsze (przedstawione jest tłumaczenie oryginalnych terminów w języku angielskim).

International Association of Assessing Officers (2018):

Mający u podstaw metody matematyczne program komputerowy, który jest przez analityków rynku stosowany w celu oszacowania wartości rynkowej, poprzez analizę wcześniej zebranych danych dotyczących lokalizacji, warunków rynkowych i charakterystyk nieruchomości. Istotną cechą modeli AVM jest to, że są efektem modelowania matematycznego. Wiarygodność tych modeli zależy od jakości danych oraz umiejętności osób tworzących modele. Modele te powinny być tworzone przy udziale wykwalifikowanych analityków rynku nieruchomości, którzy korzystają z danych oraz z symulacji aktywności rynkowej.



Są one projektowane w celu wyceny nieruchomości w wybranych momentach czasowych.

TEGOVA [w European Valuation Standard (2017)]:

AVM są to programy komputerowe mające u podstaw modele statystyczne, w celu określenia wartości nieruchomości. Modele te, stanowiące podstawy algorytmu, wykorzystują dane z bazy zawierających dane i cechy nieruchomości.

The RICS AVM Standards Working Group (2013):

AVM stosują jedną lub wiele metod matematycznych do estymacji wartości konkretnej nieruchomości w konkretnym momencie, ponadto określają poziom dokładności tej estymowanej wartości, dzieje się to bez interwencji człowieka w trakcie działania programu.

Główna cecha systemów automatycznej wyceny nieruchomości, wynikająca z powyższych określeń, to fakt, że ich podstawą są metody statystyczne wykorzystujące bazy danych, co pozwala na zakwalifikowanie ich do metod analizy statystycznej rynku.

Dokument „Standard on Automated Valuation Models (AVMs)”, opublikowany w lipcu 2018 r. przez International Association of Assessing Officers, wskazał na standardy, które powinny spełniać metody i systemy automatycznej wyceny nieruchomości.

Metody automatycznej wyceny nieruchomości, o których tu mowa, to zaawansowane metody statystyczne. Zalicza się je potocznie do metod sztucznej inteligencji (*artificial intelligence*), a ściślej, do ważnej podgrupy tych metod, zwanych metodami uczenia maszynowego (*machine learning*).

Obecnie przyjęto do metod uczenia maszynowego zaliczać również część klasycznych metod statystycznych, które znane są od wielu lat, jednak dawniej ich stosowalność (z uwagi na niewielką moc obliczeniową ówczesnych komputerów) była ograniczona do niewielkich zbiorów danych (np. kilkaset obserwacji). Obecnie można je stosować w przypadku bardzo dużych zbiorów obserwacji.

Analiza metod zaliczanych do uczenia maszynowego pozwala na ich podział na cztery podstawowe grupy:

- ▶ metody dyskryminacji;
- ▶ metody klasyfikacji;

- ▶ metody redukcji wymiaru;
- ▶ metody regresyjne.

Przy tym niektóre metody (np. sztuczne sieci neuronowe) mogą być zaliczone do więcej niż jednej grupy, gdyż mogą realizować więcej niż jedno zadanie.

Powyższa terminologia, dotycząca tych czterech grup, w języku polskim jest tradycyjnie przyjęta od wielu lat, jednak w niektórych przypadkach polski zwrot nie jest dosłownym tłumaczeniem zwrotu angielskiego.

Metody dyskryminacji (uczenie z nauczycielem, uczenie nadzorowane, *discrimination, classification*) polegają na podziale zbioru obserwacji (danych) na dwie lub więcej znanych klas, poprzez wyznaczenie jednej (lub więcej) funkcji dzielących (dyskryminujących). Na podstawie tych funkcji można następnie dokonać przydziału nowych obserwacji do jednej z wyznaczonych klas.

Metody klasyfikacji (uczenie bez nauczyciela, uczenie nienadzorowane, *classification, clustering*) polegają na podziale zbioru obserwacji (danych) na dwie lub więcej nieznanych klas. W tym przypadku nie następuje przydział nowych obserwacji do wyznaczonych klas, lecz powtórny podział na klasy takiego powiększonego zbioru obserwacji.

Metody redukcji wymiaru (*dimensionality reduction*) polegają na przekształceniu wielu zmiennych na ich zdecydowanie mniejszą liczbę (to jest właśnie redukcja wymiaru), głównie w celu bardziej przejrzystej interpretacji. Jeśli ten zredukowany wymiar wynosi 2 (ewentualnie 3), wtedy istnieje możliwość wizualizacji danych w przestrzeni dwuwymiarowej (ewentualnie trójwymiarowej).

Metody regresyjne (*regression*) polegają na wyznaczeniu funkcji, która przedstawia zależność jednej zmiennej (tzw. objaśnianej) od zbioru innych zmiennych (tzw. objaśniających).

Z powyższej listy grup metod wynika, że w procesie wyceny nieruchomości przydatne mogą być metody regresyjne. Pozostałe grupy metod mogą mieć znaczenie jedynie pomocnicze, np. przy podziale nieruchomości na klasy podobne, ewentualnie przy redukcji liczby zmiennych opisujących nieruchomość.

Sposób funkcjonowania metod wybranych do testowania w projekcie opisany jest w dalszej części rapor-



tu. Ogólnie rzecz traktując, należy wskazać, że metody te funkcjonują w dwóch etapach:

- ▶ konstrukcja funkcji na podstawie danych zbioru uczącego;
- ▶ weryfikacja funkcji na podstawie zbioru testowego.

Przy tym dla obu zbiorów danych znane są (z bazy danych) wartości nieruchomości. Dane zbioru uczącego umożliwiają konstrukcję funkcji, zaś dane zbioru testowego weryfikują jakość tej funkcji na innym zbiorze danych.

Po skonstruowaniu i zweryfikowaniu tej funkcji jest ona stosowana do wyceny nowych nieruchomości, których jeszcze nie ma w bazie danych.

1.4 Systemy automatycznej wyceny nieruchomości

Na świecie funkcjonuje już pewna liczba systemów automatycznej wyceny nieruchomości. Systemy te stosują metody statystyczne, w tym metody uczenia maszynowego, wykorzystując model (1) lub (2). Można je zatem zaliczyć do metod wyceny zaliczanych do podejścia porównawczego, ściślej do metod analizy statystycznej rynku. Niektóre systemy mają u podstaw inne, często stosowane, np. w Zjednoczonym Królestwie, podejście powtarzalnych transakcji sprzedaży (*repeated sales*). Systemy automatycznej wyceny nieruchomości powstawały najpierw w Stanach Zjednoczonych (od roku 1981), następnie w Zjednoczonym Królestwie (w latach dziewięćdziesiątych ubiegłego stulecia).

Część systemów stosowanych w Europie jest zrzeszona w European Automated Valuation Model Alliance. Są to: Calcasa (Niderlandy), CRIF (Włochy), Eiendomsverdi (Norwegia), Hometrack (Zjednoczone Królestwo), on-geo (Niemcy), on-geo (Austria), Tinsa (Hiszpania), Värderingsdata (Szwecja).

W Stanach Zjednoczonych dwa najbardziej znane systemy są zarządzane przez firmę Zillow oraz firmę HouseCanary. System Zillow wyznacza wartości domów, w tej chwili w bazie jest 110 milionów domów w USA, rozpoczęto również wycenę domów w Kanadzie. Podając adres domu, można od razu uzyskać wycenę, wartość ta nazywana jest Zestimate. System House Canary także wyznacza wartości domów. Z kolei w Zjednoczonym Królestwie dwa najbardziej znane systemy to Rightmore i Hometrack.

Od pewnego czasu toczy się dyskusja na temat zastąpienia wycen nieruchomości mieszkaniowych opracowanych przez rzeczoznawców majątkowych wycenami, które dokonuje automatyczny system wyceny nieruchomości. Jak się wydaje, są trzy kluczowe kwestie związane z przydatnością praktyczną systemów automatycznej wyceny nieruchomości:

1. Odpowiedzialność instytucji, będącej operatorem systemu, jeśli system ma być stosowany do operacji, które z powodów regulacyjnych wymagają wyceny.
2. Problem wystarczającej jakości wykorzystanych danych.
3. Przejrzystość algorytmów stosowanych w automatycznych systemach wyceny, po to, aby można było interpretować wyniki; jest to tzw. problem ryzyka modelu wyceny.



Dostępne bazy danych



2.1 Identyfikacja dostępnych baz danych

Wskazywaliśmy już powyżej, że skuteczność statystycznych metod wyceny nieruchomości mieszkaniowych (w tym metod wykorzystywanych w automatycznej wycenie) zależy m.in. od dwóch determinant, którymi są:

- ▶ dobór nieruchomości podobnych wykorzystanych do estymacji funkcji wyceny;
- ▶ wybór zmiennych charakteryzujących nieruchomość i takich, że ich wartości wpływają na wartość nieruchomości.

Wynika z tego, że bardzo ważną rolę odgrywają dane o nieruchomościach, zawarte w bazach danych. Przy tym dane te powinny być stale uzupełniane, gdyż przy dużej dynamice rynku transakcje historyczne, często nawet sprzed sześciu miesięcy, mogą istotnie zakłócać aktualną wartość nieruchomości.

Do znaczących ograniczeń baz danych o nieruchomościach należą w szczególności:

- ▶ brak możliwości wykonywania fizycznych inspekcji nieruchomości i co się z tym wiąże, konieczność przyjęcia założeń odnośnie takich cech, które nie są dostępne bezpośrednio z rynku, jak np. stan lokalu;
- ▶ brak informacji o rzeczywistej powierzchni użytkowej lokalu mieszkalnego; w przypadku wielu aktów notarialnych powierzchnia lokalu podawana jest wraz z powierzchnią komórki lokatorskiej, która nie jest powierzchniowo scharakteryzowana;
- ▶ brak informacji o cenie lokalu bez ujęcia miejsca postojowego; oznacza to, że w przypadku niektórych transakcji pojawia się cena łączna, obejmująca zarówno cenę lokalu, jak i miejsca postojowego w garażu podziemnym, o ile to miejsce nie jest wydzielone odrębną księgą wieczystą;
- ▶ brak informacji o cechach charakteryzujących mieszkania w dużych osiedlach, gdzie poszczególne lokale mogą znacząco różnić się np. z uwagi na widok z okna.

Przeprowadzona w projekcie analiza wskazała na następujące potencjalne bazy danych, z których można byłoby skorzystać:

Rejestr cen i wartości nieruchomości – rejestr publiczny prowadzony przez starostę, który zawiera dane dotyczące cen nieruchomości określonych w aktach notarialnych oraz wartości nieruchomości określonych przez rzeczoznawców majątkowych w operatach szacunkowych, których wyciągi przekazywane są do ewidencji gruntów i budynków. Oprogramowanie nie jest zestandaryzowane w skali kraju, jednakże są na rynku firmy zajmujące się standaryzacją danych otrzymanych z rejestru.

Elektroniczne księgi wieczyste – elektroniczna forma księgi wieczyste, opisującej stan prawny pojedynczej nieruchomości. Dostęp do niej jest bezpłatny, jednak nie ma możliwości wyszukania numeru księgi po adresie. Istnieją płatne portale internetowe, pozwalające na uzyskanie takich danych, jednakże nie posiadają w bazie wszystkich lokalizacji.

Bazy rzeczoznawców majątkowych – elektroniczna forma rejestru cen i wartości nieruchomości, zawierająca dane pochodzące z aktów notarialnych i wycen rzeczoznawców, stanowiąca prywatny zbiór transakcji prowadzony przez nieformalnie zrzeszone grupy rzeczoznawców majątkowych.

System Analiz i Monitorowania Rynku Obrotu Nieruchomościami (AMRON) – ogólnopolska baza danych o cenach i wartościach nieruchomości, wyposażona w funkcjonalności analityczne i statystyczne. Baza danych AMRON gromadzi informacje dotyczące charakterystyki nieruchomości, ich lokalizacji oraz informacje dotyczące cen transakcyjnych pochodzących z aktów notarialnych, wartości z wycen, ofert lub innych wiarygodnych źródeł. Pełen opis każdego typu nieruchomości zawiera średnio ok. 30 cech (zmiennych, atrybutów), dzięki czemu możliwe jest przeprowadzanie szczegółowych analiz rynku. Dodatkowo nieruchomości zarejestrowane w bazie AMRON mogą być flagowane i opisywane jako nieruchomości „podejrzane” (czyli takie, których stan faktyczny nie jest zgodny ze stanem opisanym w dokumentach, będące przedmiotem podejrzanych transakcji lub takich, których dokumenty budzą podejrzenia).

Cenatorium – oferuje aktualizowaną na bieżąco bazę cen o transakcjach na rynku nieruchomości, powstaje dzięki danym pozyskanym ze zróżnicowanych źródeł publicznych i opiera się na faktycznych cenach zakupu/sprzedaży nieruchomości.

Ewidencja podatkowa nieruchomości – zawiera dane o podatnikach i przedmiotach opodatkowania. Opiera się na danych wynikających z informacji i deklaracji składanych przez podatników, danych zawartych



w księgach wieczystych, w ewidencji gruntów i budynków oraz innych ewidencjach i rejestrach, w tym prowadzonych przez organy administracji publicznej.

Cykliczne materiały analityczne NBP – opracowanie w sposób syntetyczny opisuje najważniejsze zjawiska, jakie mają miejsce na rynku nieruchomości mieszkaniowych i komercyjnych w największych miastach w Polsce. Zawiera w szczególności informacje o cenach mieszkań i sytuacji na rynku nieruchomości mieszkaniowych i komercyjnych w Polsce, jak również bazę cen nieruchomości mieszkaniowych. Informacja ukazuje się w cyklu kwartalnym. Ponadto NBP prezentuje coroczny raport o sytuacji na rynkach nieruchomości mieszkaniowych i komercyjnych w kraju. Celem raportu jest dostarczenie zainteresowanym podmiotom gospodarczym, w tym także uczestnikom rynku nieruchomości, możliwie kompletnej, rzetelnej i obiektywnej informacji na temat sytuacji na rynku nieruchomości mieszkaniowych i komercyjnych w Polsce. Raport w głównej mierze koncentruje się na zjawiskach, które miały miejsce w danym roku i oddziałują bezpośrednio na procesy bieżące.

Geoportal – portal internetowy zapewniający dostęp do usług danych przestrzennych, tj. do usług dotyczących danych przestrzennych i będących operacjami, które mogą być wykonywane przy użyciu oprogramowania komputerowego na danych zawartych w zbiorach danych przestrzennych lub na powiązanych z nimi metadanych. Geoportal umożliwia w szczególności korzystanie z usług wyszukiwania, przeglądania, pobierania i przekształcania danych przestrzennych. Spełnia on istotną rolę w infrastrukturach danych przestrzennych (infrastrukturach informacji przestrzennej). Geoportal w szczególności zawiera dane identyfikacyjne nieruchomości, takie jak numery działek, arkusze map, obręby, powierzchnie oraz lokalizację ogólną (województwo, powiat, gmina). Pozwala również na dokonywanie pomiarów wybranych fragmentów mapy, co pomaga w ustaleniu np. powierzchni zabudowy działki. Ponadto wśród zaimplementowanych baz danych wymienić można m.in. dane o charakterze katastralnym, bazę danych ogólnie geograficznych, bazę danych obiektów topograficznych, ortofotomapy, rastry map topograficznych, rastry map tematycznych, Państwowy Rejestr Granic, Państwowy Rejestr Nazw Geograficznych, Numeryczny Model Terenu, metadane zbiorów i usług danych przestrzennych.

System Informacji Przestrzennej – zawiera informacje o przestrzeni, obiektach przestrzennych i relacjach między nimi zachodzących. W szczególności informacje te mogą dotyczyć planów miejscowych, studium zagospodarowania przestrzennego, danych

demograficznych, usytuowania wybranych punktów POI, map akustycznych, map wysokościowych, map rowerowych i innych. Zakres danych zależy od decyzji gminy prowadzącej SIP o stworzeniu i udostępnieniu wybranych informacji przestrzennych.

Raporty JLL – raporty, analizy, i trendy dotyczące rynku nieruchomości (również mieszkaniowego), przygotowane przez firmę doradczą, świadczącą profesjonalne usługi na rynku nieruchomości.

Badania i analizy rynku grupy Emerson – płatne raporty przygotowane przez analityków firmy, prezentujące w szczególności analizy oparte na monitorowaniu rynku pierwotnego i wtórnego, danych podaży i popytu, zarówno cen ofertowych, jak i transakcyjnych.

Serwisy ogłoszeń sprzedaży i wynajmu nieruchomości (m.in. otodom.pl, domiporta.pl, morizon.pl, domy.pl, trovit.pl, rynekpierwotny.pl, szybko.pl i inne) – dane dotyczą wyłącznie wartości ofertowych, występujących zarówno na rynku pierwotnym, jak i wtórnym. Istnieją firmy agregujące dane pochodzące z wyżej wymienionych portali w postaci elektronicznej, w formacie xls bądź innym, umożliwiającym dokonanie dalszych analiz.

Google Maps – serwis internetowy umożliwiający wyszukiwanie obiektów, oglądanie map, zdjęć lotniczych powierzchni Ziemi, 360-stopniowe panoramiczne widoki z poziomu ulic (Street view), natężenie ruchu ulicznego w czasie rzeczywistym, planowanie tras podróży samochodem, transportem publicznym, rowerem, pieszo lub samolotem.

Airly.eu – dane przestrzenne o zanieczyszczeniu powietrza w Polsce.

2.2 Określenie zmiennych wejściowych i weryfikacja ich dostępności w bazach danych

Ważnym etapem tworzenia systemu automatycznej wyceny nieruchomości jest określenie zmiennych wejściowych (charakteryzujących nieruchomość), a następnie weryfikacja ich dostępności w zidentyfikowanych bazach danych. W projekcie wykonano w tym zakresie dwa zadania:

- ▶ sformułowanie zestawu zmiennych wejściowych – na podstawie wiedzy merytorycznej wynikającej z analizy fundamentalnej nieruchomości oraz wyników badań autorów różnych publikacji;



- ▶ konstrukcja zerojedynkowej macierzy dostępności, poprzez wskazanie baz danych, w których dostępne są wartości danej zmiennej wejściowej (w macierzy tę sytuację oznacza się liczbą 1) lub brak jest jakiegokolwiek wartości tej zmiennej (w macierzy tę sytuację oznacza się liczbą 0).

Jeśli chodzi o zmienne wyjściowe, to dokonany został ich podział na dwa rodzaje zmiennych (w odniesieniu do nieruchomości mieszkaniowych):

- ▶ zmienne charakteryzujące samo mieszkanie;
- ▶ zmienne charakteryzujące najbliższą okolicę mieszkania.

Lista potencjalnych zmiennych charakteryzujących samo mieszkanie jest następująca (wraz z podaniem możliwych wariantów wartości):

- ▶ Lokalizacja
 - adres
 - dane GPS
- ▶ Powierzchnia użytkowa
- ▶ Piętro
 - pierwsze, drugie, trzecie lub wyższe, ostatnie
 - które i z ilu
- ▶ Liczba pokoi
- ▶ Liczba łazienek
- ▶ Osobna jasna kuchnia
 - tak/nie
- ▶ Balkon/taras
 - tak/nie oraz powierzchnia, jeśli znacząco odbiega od standardów rynkowych
- ▶ Standard lokalu
 - wysoki, dobry, przeciętny, do wykończenia/do remontu
 - pod klucz, do odświeżenia, do remontu, do generalnego remontu, stan surowy
- ▶ Winda
 - tak/nie
- ▶ Piwnica/komórka lokatorska
 - tak/nie oraz powierzchnia, jeśli znana bądź znacząco przekraczająca standardy rynkowe
- ▶ Rodzaj miejsca parkingowego
 - miejsce postojowe/garaż w budynku
- ▶ Rodzaj budynku
 - blok, kamienica, apartamentowiec, wieżowiec (do 20 pięter), inny
- ▶ Rok budowy
 - okres przedwojenny, lata 40., lata 50., lata 60., lata 70., lata 80., lata 90., 2000–2010, po 2010 r.

- do 1945, od 1946 do 1970, od 1971 do 1991, od 1992 do 2001, od 2002

- ▶ Rynek
 - pierwotny/wtórny
- ▶ Ceny transakcyjne
- ▶ Ceny ofertowe
- ▶ Data transakcji
- ▶ Prawo do lokalu
 - użytkowanie wieczyste/własność/spółdzielcze własnościowe

Lista potencjalnych zmiennych charakteryzujących najbliższą okolicę mieszkania jest następująca (wraz z podaniem możliwych wariantów wartości):

- ▶ Skomunikowanie z innymi dzielnicami
- ▶ Obecność szkół
- ▶ Obecność sklepów
- ▶ Odległość do najbliższego punktu danego typu (miejsca użyteczności publicznej, źródło: Open Street Map):
 - metro, autobus, tramwaj, PKP
 - szkoła, przedszkole,
 - szpital, przychodnia
 - galeria handlowa
 - tereny zieleni
- ▶ Czas dojazdu
 - do centrum miasta
 - do wybranych istotnych punktów miasta
- ▶ Bezpieczeństwo (dane o przestępstwach i wykroczeniach pozyskane ze źródeł publicznych)
 - wskaźnik poziomu bezpieczeństwa okolicy – wyliczany na bazie zdarzeń zaistniałych w promieniu X km, dlatego wartość wskaźnika może różnić się od średniego poziomu bezpieczeństwa dla dzielnicy
 - średni poziom bezpieczeństwa w dzielnicy – wyższa wartość wskaźnika oznacza bardziej bezpieczną dzielnicę, czyli mniej zdarzeń lub zdarzenia sklasyfikowane jako mniej istotne z punktu widzenia bezpieczeństwa
- ▶ Zanieczyszczenie powietrza
 - średnie miesięczne stężenie pyłu zawieszonego dla najbliższej stacji pomiarowej ($\mu\text{g}/\text{m}^3$)
 - wskaźnik poziomu zanieczyszczenia powietrza dla dzielnicy, na podstawie uśrednionych rocznych pomiarów stężenia pyłów zawieszonych PM10 oraz PM2,5; im wyższa wartość tego wskaźnika, tym lepsza jest jakość powietrza w danej okolicy

Kolejnym etapem realizacji projektu była konstrukcja zerojedynkowej macierzy dostępności, poprzez wska-



zanie baz danych, w których dostępne są wartości danej zmiennej wejściowej. Wyniki analizy przedstawia **tabela 1**, zawierająca tę macierz dostępności. Dostępność danej zmiennej w wybranej bazie danych oznaczana jest liczbą 1, zaś brak tej dostępności liczbą 0.

Przy wyborze bazy danych o nieruchomościach na potrzeby projektu istotne były następujące kryteria:

- ▶ kompletność danych;
- ▶ dostępność bazy (również pod kątem finansowym);
- ▶ reprezentatywność danych;
- ▶ ciągłość w aktualizacji wkładu informacyjnego;
- ▶ liczba i istotność uwzględnianych czynników;

- ▶ błędy mechaniczne w danych;
- ▶ doświadczenie podmiotu dostarczającego bazę;
- ▶ łatwość obsługi bazy poprzez eksport danych do systemów umożliwiających przeprowadzenie analiz statystycznych i wycen;
- ▶ spójność formatu danych (integralność semantyczna).

Kryterium kompletności danych oraz liczby i istotności uwzględnianych czynników pozwoliło na szybką eliminację tych baz, które dostarczają wybiórczych informacji o rynku nieruchomości (np. Airly.eu, Google Maps. Geoportal) i stanowić mogą jedynie uzupełnienie przeprowadzanych analiz, nie zaś ich istotny trzon. W dalszej kolejności wyeliminowane zostały

Tabela 1. Dostępność danych wejściowych w bazach danych

	Lokalizacja	Powierzchnia użytkowa	Piętro	Liczba kondygnacji	Liczba pokoi	Liczba łazienek	Osobna jasna kuchnia	Balkon/taras	Standard lokalu	Winda	Piwnica/komórka lokatorska	Rodzaj miejsca parkingowego	Rodzaj budynku	Rok budowy	Rynek	Skomunikowanie z innymi dzielnicami	Obecność szkół	Obecność sklepów	Odległość od najbliższego punktu danego typu	Czas dojazdu	Bezpieczeństwo	Zanieczyszczenie powietrza	Ceny transakcyjne	Ceny ofertowe	Data transakcji	Prawo do lokalu	
Rejestr cen i wartości nieruchomości	1	1	1	0	1	1	0	0/1	0/1	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	1	1	
Elektroniczne księgi wieczyste	1	1	0/1	0	0/1	0/1	0	0/1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	
Bazy rzeczoznawców majątkowych	1	1	1	0	1	1	0	0/1	0/1	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	1	1	
Cenatorium	1	1	1	1	1	1	0/1	0/1	0/1	0/1	0/1	1	1	0/1	1	0	0	0	0	0	0	0	1	0	1	1	
Ewidencja podatkowa nieruchomości	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	
Cykliczne materiały analityczne NBP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	
Geoportal	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	
System Informacji Przestrzennej	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	0	1	0/1	0	0	0	0	
System Analiz i Monitorowania Rynku Obrotu Nieruchomościami – AMRON	1	1	1	0	1	1	0	0/1	0/1	0	1	1	0/1	0/1	1	0	0	0	0	0	0	0	1	0	1	1	
Raporty branżowe – JLL	0/1	0/1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0/1	0/1	0/1	0	
Badania i analizy rynku grupy Emmerson	0/1	0/1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0/1	0/1	0	
Serwisy ogłoszeń sprzedaży i wynajmu nieruchomości	1	1	1	0/1	1	1	0/1	0/1	1	0/1	0/1	1	0/1	0/1	1	0/1	0/1	0/1	0/1	0/1	0	0/1	0	0	1	0	1
Google Maps	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	
Airly.eu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	

bazy tworzone na podstawie danych dostarczanych przez inne podmioty kompletujące dane pierwotne (cykliczne materiały analityczne NBP). Z uwagi na fakt, iż wyceny nieruchomości bazują na danych transakcyjnych, wykluczone zostały również bazy dostarczające wyłącznie ceny ofertowe.

Ostatecznie rozważane były bazy: System Analiz i Monitorowania Rynku Obrotu Nieruchomościami (AMRON), rejestr cen i wartości nieruchomości, bazy rzeczoznawców majątkowych oraz Cenatorium.

Z uwagi na dynamikę rynku nieruchomości niezwykle istotne jest ujmowanie przy wycenie na rynku pierwotnym cen w okresie decyzji o zakupie i podjęcia pierwszych kroków (zawarcie umowy kredytowej). Jedynie baza AMRON z pozostałych analizowanych baz danych posiada dane dotyczące historii nieruchomości, tj. historii i zdarzeń związanych z nieruchomością oraz opis zmian fizycznych samej nieruchomości. Pozostałe bazy opierają się na momencie zawarcia ostatecznego aktu notarialnego, którego data może różnić się od daty podjęcia decyzji o zakupie nawet o kilkanaście miesięcy.

Bazę AMRON wyróżnia także reprezentatywność danych (rozkład terytorialny zasobów bazy odpowiada głębokości rynków lokalnych oraz aktywności kredytowej polskiego sektora bankowego) oraz długie doświadczenie w funkcjonowaniu na rynku polskim.

Ponadto należy zauważyć, że baza ta jest tworzona z danych dostarczanych przez banki, które będą głównymi interesariuszami systemu automatycznej wyceny nieruchomości.

Podane argumenty uzasadniają wybór bazy AMRON na potrzeby badań przeprowadzanych w projekcie.

2.3 Metody wyodrębniania zbiorów nieruchomości podobnych

Zarówno polskie regulacje, jak i dotychczasowa praktyka oraz badania naukowe wskazują, że wycena nieruchomości mieszkaniowych ma u podstaw dane dotyczące nieruchomości podobnych. Oznacza to, że etapem poprzedzającym wycenę z zastosowaniem funkcji (1) lub (2) może być wyodrębnienie zbiorów nieruchomości podobnych. Funkcja wyceny (1) lub (2) jest w takiej sytuacji inna w przypadku każdego zbioru nieruchomości podobnych, zaś dla danej wycenianej nieruchomości wybierana jest oczywiście funkcja odpowiadająca zbiorowi, do którego należy ta nieruchomość.

W celu wyodrębnienia zbioru nieruchomości podobnych stosowane są metody klasyfikacji. Jak już wskazywaliśmy, polegają one na podziale zbioru obiektów, opisywanych przez pewną liczbę zmiennych, w tym wypadku charakterystyk nieruchomości, na dwie lub więcej nieznanych wcześniej klas. Od co najmniej kilkudziesięciu lat takie metody są proponowane w literaturze, a od dawna część z nich jest stosowana w praktyce. Opis metod klasyfikacji przedstawiony jest np. w książce: Everitt B., Landau S., Leese M., Stahl D. (2011).

Ogół najczęściej stosowanych w praktyce metod klasyfikacji można podzielić na dwie grupy:

- ▶ hierarchiczne metody grupowania;
- ▶ metody optymalizacyjne.

W obu grupach metod podobieństwo obiektów (w tym wypadku nieruchomości) jest określane za pomocą funkcji odległości, w myśl zasady: im bliższa odległość między obiektami, tym większe ich podobieństwo. Odległość między obiektami charakteryzującymi się wieloma zmiennymi może być wyznaczona np. za pomocą odległości euklidesowej, według wzoru:

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

3

gdzie:

d_{ij} – odległość między obiektem i oraz obiektem j ;
 x_{ik} – wartość zmiennej k dla obiektu i .

W hierarchicznych metodach grupowania proces klasyfikacji przebiega w następujący sposób:

- ▶ punktem wyjścia są jednoelementowe zbiory obiektów (w tym wypadku nieruchomości), tzn. każdy obiekt stanowi jeden zbiór;
- ▶ w pierwszym etapie następuje połączenie dwóch najbardziej podobnych jednoelementowych zbiorów, czyli takich, dla których odległość między nimi jest najmniejsza; w ten sposób liczba zbiorów zmniejsza się o 1;
- ▶ w każdym kolejnym etapie powtarza się ten proces, poprzez łączenie dwóch najbliższych zbiorów i w ten sposób w każdym etapie liczba zbiorów zmniejsza się o 1;



- ▶ proces ten kontynuuje się do momentu, gdy otrzymuje się jeden zbiór zawierający wszystkie obiekty.

W ten sposób otrzymana jest sekwencja klasyfikacji, spośród których wybiera jest ta, w której liczba zbiorów równa jest zadanej wcześniej liczbie lub ta, w której jest maksymalna wartość wskaźnika jakości klasyfikacji.

Metody optymalizacyjne polegają na wyborze takiej klasyfikacji zbioru obiektów na zadaną liczbę klas, dla której to klasyfikacji:

- ▶ odległości między obiektami należącymi do tych samych klas są jak najmniejsze;
- ▶ odległości między obiektami należącymi do różnych klas są jak największe.

Jedną z najbardziej popularnych metod optymalizacyjnych jest metoda *k-średnich* (*k-means*). Była ona stosowana w badaniach tego projektu. Metoda ta przebiega według następującego schematu:

- ▶ punktem wyjścia jest klasyfikacja początkowa, która może być zadana losowo;
- ▶ w każdym etapie:
 - a) wyznacza się tzw. środki ciężkości klas, są to „sztuczne” obiekty, dla których wartości zmien-

nych są określone jako średnie arytmetyczne wartości zmiennych dla obiektów należących do tej klasy;

- b) wyznacza się odległości każdego obiektu od środków ciężkości wszystkich klas;
- c) dokonuje się reklasyfikacji obiektów poprzez przydzielenie każdego do tej klasy, w przypadku której odległość obiektu od środka ciężkości tej klasy jest najmniejsza;

- ▶ opisane działania są kontynuowane w kolejnych etapach, do uzyskania klasyfikacji, która się nie zmienia w kolejnym etapie.

W przypadku zastosowania metod klasyfikacji do wyznaczenia zbioru nieruchomości podobnych są jeszcze cząstkowe zagadnienia, które powinny być rozstrzygnięte:

- ▶ wybór odległości do pomiaru podobieństwa nieruchomości (zwłaszcza w przypadku zmiennych mierzonych na różnych skalach);
- ▶ wybór liczby klas; może być rozstrzygnięty na podstawie wiedzy eksperckiej lub przez zastosowanie miernika jakości klasyfikacji;
- ▶ wybór samej metody klasyfikacji.



Zbiór potencjalnych metod wyceny





3.1 Opis zbioru potencjalnych metod wyceny

Jak wynika z poprzednich rozważań, potencjalne metody wyceny, które mogą być zastosowane w systemie automatycznej wyceny nieruchomości, to metody regresyjne. Polegają one na wyborze postaci funkcji we wzorze (1) lub (2), a następnie oszacowaniu parametrów tej funkcji na podstawie dostępnych danych dotyczących zmiennej objaśnianej (w tym przypadku cen nieruchomości lub logarytmów naturalnych cen nieruchomości) oraz wartości zmiennych objaśniających (charakteryzujących nieruchomości).

Na potrzeby testowania w projekcie zostały wybrane następujące metody:

- ▶ Metoda regresji liniowej w odniesieniu do wartości nieruchomości [wzór (1)];
- ▶ Metoda regresji liniowej w odniesieniu do logarytmu wartości nieruchomości [wzór (2)];
- ▶ Metoda regresji LASSO;
- ▶ Metoda liniowej regresji bayesowskiej;
- ▶ Metoda lasów losowych;
- ▶ Metoda wektorów nośnych (SVM).

Wybór na potrzeby testowania podyktowany był prostotą i popularnością tych metod. W tworzonego systemu automatycznej wyceny jest oczywiście możliwość oprogramowania większej liczby metod, aby na potrzeby wyceny określonej klasy nieruchomości wybierana była optymalna – w danym momencie – metoda.

Najprostszy przypadek to model regresji liniowej w postaci:

$$P = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \epsilon$$

lub:

$$\ln P = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \epsilon$$

Gdzie:

$\beta_0, \beta_1, \beta_2, \dots, \beta_m$ – parametry modelu liniowego.

Parametry modeli (1) lub (2) – podobnie modeli (4) lub (5) – estymuje się na podstawie danych historycznych. W dalszej części stosowane są następujące oznaczenia:

n – liczba nieruchomości w zbiorze danych,

$p_1, p_2, \dots, p_n$ – ceny nieruchomości,

$x_{1j}, x_{2j}, \dots, x_{mj}$ – wartości zmiennej.

Regresja liniowa

Najpopularniejszy jest oczywiście model regresji liniowej. Estymacja parametrów w tym wypadku przeprowadzana jest z zastosowaniem metody najmniejszych kwadratów. Polega ona na minimalizacji – ze względu na wartości parametrów – funkcji:

$$L(\beta_0, \beta_1, \beta_2, \dots, \beta_m) = \sum_{i=1}^n (p_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im}))^2$$

Przy tym równanie (6) może być stosowane również do logarytmów cen nieruchomości [jak w modelu (2)].

Regresja LASSO

Metoda regresji LASSO (*Least Absolute Shrinkage and Selection Operator*) w porównaniu z klasyczną metodą regresji charakteryzuje się dwoma cechami. Po pierwsze, ogranicza negatywny wpływ współliniowości zmiennych objaśniających na wyniki estymacji. Po drugie, w przypadku nieistotnych zmiennych objaśniających wartości współczynników stojących przy nich są równe 0.

Estymacja parametrów w metodzie regresji LASSO polega na minimalizacji – ze względu na wartości parametrów – funkcji:

$$L(\beta_0, \beta_1, \beta_2, \dots, \beta_m) = \sum_{i=1}^n (p_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im}))^2 + \lambda \sum_{j=1}^m |\beta_j|$$

Liniowa regresja bayesowska

W klasycznej regresji liniowej estymacja jest przeprowadzana metodą najmniejszych kwadratów. W przypadku rozkładu normalnego wartości zmiennych daje ona taki sam wynik jak metoda największej wiarygodności, która formalnie polega na



maksymalizacji funkcji wiarygodności wynikającej z funkcji gęstości. Przy założeniu rozkładu normalnego funkcja wiarygodności dla regresji liniowej ma postać:

$$L(\beta_0, \beta_1, \beta_2, \dots, \beta_m) = c \frac{1}{2\sigma^2} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (p_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im}))^2 \right)$$

8

W przypadku regresji bayesowskiej, oprócz założenia o rozkładzie zmiennych, czyni się założenie o rozkładzie parametrów, którymi są współczynniki przy zmiennych oraz wariancja składnika losowego. Jest to tzw. rozkład *a priori*. Połączenie tego rozkładu z rozkładem zmiennych daje (na mocy twierdzenia Bayesa) tzw. rozkład *a posteriori*. W odniesieniu do tego rozkładu stosowana jest następnie funkcja wiarygodności, której maksymalizacja prowadzi do ocen parametrów, czyli m.in. współczynników przy zmiennych.

Metoda lasów losowych

Metoda lasów losowych (*Random Forest Method*) ma u podstaw regułę drzew decyzyjnych. Polega ona na tym, że sekwencyjnie dla każdej zmiennej podejmuje się decyzję co do przydzielenia wartości tej zmiennej do odpowiedniej klasy wartości. Metoda lasów losowych polega na konstrukcji zbioru drzew decyzyjnych i wyboru tego drzewa, które daje najlepsze dopasowanie funkcji. Przy tym ten proces przebiega wielokrotnie do momentu otrzymania najlepszej funkcji w przypadku zbioru testowego.

Metoda SVM

Metoda SVM, zwana metodą wektorów nośnych (*Support Vector Machines*) polega na wyznaczeniu zależności między zmiennymi na podstawie znanej zależności dla przypadków zbioru uczącego, a następnie dopasowaniu tej funkcji w odniesieniu do zbioru testowego, po to, aby tak otrzymana funkcja była zastosowana do wyceny. Przy tym wyznaczanie funkcji odbywa się poprzez minimalizowanie w kolejnych etapach funkcji dopasowania podobnej do funkcji wyrażonej wzorem (5).

3.2 Określenie jakości funkcjonowania metody wyceny

Ważnym elementem określenia jakości funkcjonowania danej metody analizy statystycznej rynku jest porównanie otrzymanej (oszacowanej) wartości nieruchomości z wartością rynkową. Dotyczy to:

- ▶ zbioru uczącego, to pozwala na ocenę dobroci dopasowania oszacowanego modelu (1) lub (2) do danych;
- ▶ zbioru testowego, to pozwala na ocenę jakości funkcjonowania modelu wyceny w procesie automatycznej wyceny.

Ocena ta może być dokonana za pomocą mierników dobroci dopasowania. W projekcie zastosowano pięć mierników. Przedstawione są one poniżej, oznaczenia są następujące:

n – liczba nieruchomości w zbiorze danych (zbiór uczący lub zbiór testowy),

$p_1, p_2, \dots, p_n$ – ceny nieruchomości,

$p_1^e, p_2^e, \dots, p_n^e$ – wartości nieruchomości otrzymane w procesie wyceny.

Mierniki te mierzą dobroć dopasowania bądź jakość wyceny poprzez analizę różnic między cenami a wartościami oszacowanymi w procesie wyceny.

Oczywiście w systemie jest możliwość stosowania dowolnych mierników dobroci dopasowania.

Oto lista podstawowych mierników.

1. Błąd średniokwadratowy (MSE – *Mean Squared Error*).

Dany jest jako średnia z sumy kwadratów różnic między dwoma zbiorami wartości:

$$MSE = \frac{1}{n} \sum_{i=1}^n (p_i - p_i^e)^2$$

9

2. Pierwiastek błędu średniokwadratowego (RMSE – *Root Mean Squared Error*):

Dany jest jako pierwiastek średniej z sumy kwadratów różnic między dwoma zbiorami wartości:



10

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - p_i^e)^2} = \sqrt{MSE}$$

3. Pierwiastek z procentowego błędu średniokwadratowego (RMSPE – Root Mean Squared Percentage Error):

Dany jest jako pierwiastek średniej z sumy kwadratów procentowych różnic między dwoma zbiorami wartości:

11

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{p_i - p_i^e}{p_i} \right)^2}$$

4. Średni błąd bezwzględny (MAE – Mean Absolute Error):

Dany jest jako średnia z sumy bezwzględnych różnic między dwoma zbiorami wartości:

12

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - p_i^e|$$

5. Średni bezwzględny błąd procentowy (MAPE – Mean Absolute Percentage Error):

Dany jest jako średnia z sumy bezwzględnych różnic procentowych między dwoma zbiorami wartości:

13

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{p_i - p_i^e}{p_i} \right|$$

Oczywiście w przypadku wszystkich tych mierników interpretacja jest taka sama: im niższa wartość miernika, tym lepsza jakość metody. Z punktu widzenia użytkownika końcowego ważne są mierniki (11) i (13), gdyż odnoszą się do różnic procentowych, co wydaje się najprostsze do interpretacji.

3.3 Badania w zakresie zastosowania metod analizy statystycznej rynku i możliwości ich implementacji w programie CRAN R – schemat i wyniki badań

W projekcie przeprowadzone zostały testowe badania wybranych metod analizy statystycznej rynku, mających charakter metod regresyjnych w odniesieniu do wybranych nieruchomości pochodzących z bazy AMRON. Do implementacji tych metod wybrany został program R, ściślej system CRAN R (*Comprehensive R Archive Network*). Jest to w tej chwili dominujący w praktyce niekomercyjny profesjonalny program służący do analizy danych za pomocą metod statystycznych.

Oczywiście w systemie jest możliwość stosowania dowolnego programu implementującego metody statystyczne.

W **tabeli 2** przedstawione są pakiety programu CRAN R stosowane w metodach analizy statystycznej rynku.

Tabela 2. Metody wyceny i pakiety programu CRAN R

Metoda wyceny	Pakiety programu R
Regresja liniowa	hpiR
Regresja typu LASSO	lasso2, glmnet
Regresja bayesowska	bayesm
Metoda lasów losowych	tree, rpart, randomForest, maptree
Metoda regresyjnych wektorów nośnych (SVM)	e1071, liquidSVM

Źródło: R Core Team (2020).

Badania przeprowadzone w projekcie przebiegały według następującego schematu:

- ▶ podział zbioru danych na dwie klasy: zbiór uczący (treningowy) i zbiór testowy;
- ▶ zastosowanie wybranych metod do danych zbioru uczącego w celu oszacowania modelu (1) lub (2);
- ▶ zastosowanie oszacowanego modelu w odniesieniu do danych zbioru testowego w celu uzyskania wartości nieruchomości i porównania jej z wartością rynkową, dostępną w bazie danych.

W testowaniu metod analizie podlegało sześć zbiorów danych, pozyskanych z bazy AMRON. Dotyczyły one sześciu miast. Wyeliminowano niewielką liczbę nieruchomości, w przypadku których były braki wartości niektórych zmiennych. Ilość nieruchomości, które zostały analizowane, wynosiła dla poszczególnych miast:

- ▶ Jelenia Góra – 796;
- ▶ Poznań – 7800;
- ▶ Warszawa – 13 572;
- ▶ Wrocław – 7020;
- ▶ Zamość – 318;
- ▶ Zielona Góra – 1032.

Oprócz tego wygenerowany został siódmy zbiór danych, zawierający losowo wybrane nieruchomości z tych sześciu miast, w sumie 600 z nich.

Przeprowadzono też analizę dla połączonych zbiorów danych (wszystkie sześć miast), zbiór ten zawierał dane dotyczące 30 538 nieruchomości.

Ponadto w przypadku dwóch miast (Jelenia Góra, Warszawa) dokonano klasyfikacji nieruchomości na trzy klasy nieruchomości podobnych, za pomocą metody k-średnich – analiza przeprowadzona została odrębnie dla każdej z tych trzech klas.

Oznacza to, że w sumie analizowane było 14 zbiorów nieruchomości.

W przypadku każdego miasta (również w przypadku losowego zbioru oraz zbioru zawierającego dane dotyczące wszystkich sześciu miast) dane zostały podzielone na dwa równej wielkości podzbiory. Pierwszy podzbiór to zbiór uczący, na podstawie którego został wyznaczony model wyceny. Drugi podzbiór to zbiór testowy; w jego przypadku wartości wybranych zmiennych zostały podstawione do modelu wyceny, a następnie otrzymane wartości wycenianych nieruchomości zostały porównane z cenami tych nieruchomości dostępnymi w bazie danych.

Na potrzeby testowania wybrany został **jednolity zestaw zmiennych charakteryzujących nieruchomości**, były to następujące zmienne: powierzchnia mieszkania, liczba pokoi, piętro, liczba pięter w budynku, rok budowy.

W **tabelach 3–16** przedstawione są wyniki dla poszczególnych zbiorów danych. Zawierają one **błędy procentowe** [wzory (11) i (13)] w przypadku każdej z rozważanych metod. Przy tym tabele 4–6 zawierają obliczenia dla odrębnych klas w przypadku Jeleniej Góry, a tabele 9–11 zawierają obliczenia dla odrębnych klas w przypadku Warszawy.

Tabela 3. Błędy dla zbioru testowego – Jelenia Góra

	RMSPE	MAPE
Liniowa	0,59	0,42
Lasso	0,53	0,38
SVM	0,56	0,41
Lasy losowe	0,56	0,40
Bayesowska	0,36	0,25
Liniowa (2)	0,38	0,26

Źródło: opracowanie własne.

Tabela 4. Błędy dla zbioru testowego – Jelenia Góra (klasa 1 – 67 nieruchomości)

	RMSPE	MAPE
Liniowa	0,18	0,14
Lasso	0,18	0,13
SVM	0,17	0,13
Lasy losowe	0,17	0,13
Bayesowska	0,13	0,10
Liniowa (2)	0,11	0,09

Źródło: opracowanie własne.

Tabela 5. Błędy dla zbioru testowego – Jelenia Góra (klasa 2 – 474 nieruchomości)

	RMSPE	MAPE
Liniowa	0,44	0,28
Lasso	0,43	0,26
SVM	0,49	0,29
Lasy losowe	0,48	0,28
Bayesowska	0,36	0,23
Liniowa (2)	0,34	0,23

Źródło: opracowanie własne.



Tabela 6. Błędy dla zbioru testowego
– Jelenia Góra (klasa 3 – 255 nieruchomości)

	RMSPE	MAPE
Liniowa	0,15	0,13
Lasso	0,15	0,13
SVM	0,16	0,14
Lasy losowe	0,17	0,14
Bayesowska	0,14	0,12
Liniowa (2)	0,14	0,12

Źródło: opracowanie własne.

Tabela 9. Błędy dla zbioru testowego
– Warszawa (klasa 1 – 10 942 nieruchomości)

	RMSPE	MAPE
Liniowa	0,34	0,25
Lasso	0,31	0,23
SVM	0,35	0,27
Lasy losowe	0,34	0,25
Bayesowska	0,19	0,15
Liniowa (2)	0,21	0,15

Źródło: opracowanie własne.

Tabela 7. Błędy dla zbioru testowego
– Poznań

	RMSPE	MAPE
Liniowa	0,58	0,37
Lasso	0,52	0,33
SVM	0,56	0,36
Lasy losowe	0,56	0,36
Bayesowska	0,28	0,16
Liniowa (2)	0,32	0,16

Źródło: opracowanie własne.

Tabela 10. Błędy dla zbioru testowego
– Warszawa (klasa 2 – 2367 nieruchomości)

	RMSPE	MAPE
Liniowa	0,26	0,21
Lasso	0,24	0,20
SVM	0,27	0,21
Lasy losowe	0,28	0,22
Bayesowska	0,18	0,14
Liniowa (2)	0,18	0,14

Źródło: opracowanie własne.

Tabela 8. Błędy dla zbioru testowego
– Warszawa

	RMSPE	MAPE
Liniowa	0,73	0,48
Lasso	0,64	0,43
SVM	0,66	0,44
Lasy losowe	0,69	0,43
Bayesowska	0,26	0,19
Liniowa (2)	0,29	0,18

Źródło: opracowanie własne.

Tabela 11. Błędy dla zbioru testowego
– Warszawa (klasa 3 – 163 nieruchomości)

	RMSPE	MAPE
Liniowa	0,44	0,32
Lasso	0,38	0,29
SVM	0,36	0,27
Lasy losowe	0,42	0,30
Bayesowska	0,26	0,20
Liniowa (2)	0,24	0,19

Źródło: opracowanie własne.

Tabela 12. Błędy dla zbioru testowego
– Wrocław

	RMSPE	MAPE
Liniowa	0,60	0,38
Lasso	0,55	0,34
SVM	0,58	0,37
Lasy losowe	0,58	0,36
Bayesowska	0,34	0,18
Liniowa (2)	0,35	0,18

Źródło: opracowanie własne.

Tabela 15. Błędy dla zbioru testowego zawierającego
wszystkie 6 miast

	RMSPE	MAPE
Liniowa	1,39	0,83
Lasso	1,29	0,80
SVM	1,29	0,76
Lasy losowe	1,35	0,79
Bayesowska	0,83	0,49
Liniowa (2)	0,77	0,45

Źródło: opracowanie własne.

Tabela 13. Błędy dla zbioru testowego
– Zamość

	RMSPE	MAPE
Liniowa	0,54	0,37
Lasso	0,50	0,34
SVM	0,50	0,37
Lasy losowe	0,50	0,36
Bayesowska	0,37	0,21
Liniowa (2)	0,37	0,23

Źródło: opracowanie własne.

Tabela 16. Błędy dla zbioru testowego – losowo wybrane
nieruchomości

	RMSPE	MAPE
Liniowa	1,20	0,79
Lasso	1,21	0,81
SVM	1,01	0,70
Lasy losowe	1,02	0,71
Bayesowska	0,86	0,57
Liniowa (2)	0,72	0,49

Źródło: opracowanie własne.

Tabela 14. Błędy dla zbioru testowego
– Zielona Góra

	RMSPE	MAPE
Liniowa	0,43	0,33
Lasso	0,38	0,29
SVM	0,42	0,32
Lasy losowe	0,41	0,31
Bayesowska	0,20	0,15
Liniowa (2)	0,21	0,16

Źródło: opracowanie własne.

Z kolei w **tabelach 17–20** zestawione są błędy procentowe dla wszystkich zbiorów danych i wszystkich metod. Przy tym:

- ▶ RMSPE jest to pierwiastek z procentowego błędu średniokwadratowego, czyli pierwiastek średniej z sumy kwadratów procentowych różnic między dwoma zbiorami wartości;
- ▶ MAPE jest to średni bezwzględny błąd procentowy, dany jest jako średnia z sumy bezwzględnych różnic procentowych między dwoma zbiorami wartości.

Analiza wyników zawartych w **tabelach 17–18** pozwala na wyciągnięcie następujących konkluzji:



- ▶ najgorsze wyniki uzyskano dla zbioru łącznego i zbioru losowego, co jest zrozumiałe z uwagi na niejednorodność tych zbiorów;
- ▶ zdecydowanie lepsze wyniki są dla miast niż dla łącznych zbiorów, z tym, że lepsze są w przypadku mniejszych miast (Zielona Góra, Zamość), wyjątkiem są dobre wyniki w przypadku Warszawy dla dwóch metod;
- ▶ porównanie sześciu zastosowanych metod wskazuje, że zdecydowanie najlepsze wyniki uzyskano w przypadku regresji bayesowskiej oraz regresji liniowej dla logarytmu cen;
- ▶ **błędy są wyższe niż 10%, jednak testowanie metod było dokonane bez** wcześniejszej klasyfikacji na nieruchomości podobne.

Jak widać, najważniejsze konkluzje z tych dwóch tabel dotyczą wyboru metod.

Z kolei analizy wyników zawartych w tabelach 19–20 wskazuje, że – czego można było się spodziewać – klasyfikacja na zbiory nieruchomości podobnych prowadzi do znacznej poprawy. Przy tym w klasyfikacji przyjęto liczbę klas jako 3 (przykładowo). W przypadku tworzenia systemu trzeba dokonać optymalizacji liczby klas z punktu widzenia dokładności wyników, co może być dokonane różnymi metodami – jest to jedno z zagadnień, które były zasygnalizowane w rozdziale 2.3 tego raportu.

Tabela 17. Zestawienie wartości błędu RMSPE – bez podziału na nieruchomości podobne

	Jelenia Góra	Poznań	Warszawa	Wrocław	Zamość	Zielona Góra	Łączne	Losowe
Liniowa	0,59	0,58	0,73	0,60	0,54	0,43	1,39	1,20
Lasso	0,53	0,52	0,64	0,55	0,50	0,38	1,29	1,21
SVM	0,56	0,56	0,66	0,58	0,50	0,42	1,29	1,01
Lasy losowe	0,56	0,56	0,69	0,58	0,50	0,41	1,35	1,02
Bayesowska	0,36	0,28	0,26	0,34	0,37	0,20	0,83	0,86
Liniowa (2)	0,38	0,32	0,29	0,35	0,37	0,21	0,77	0,72

Źródło: opracowanie własne.

Tabela 18. Zestawienie wartości błędu MAPE – bez podziału na nieruchomości podobne

	Jelenia Góra	Poznań	Warszawa	Wrocław	Zamość	Zielona Góra	Łączne	Losowe
Liniowa	0,42	0,37	0,48	0,38	0,37	0,33	0,83	0,79
Lasso	0,38	0,33	0,43	0,34	0,34	0,29	0,80	0,81
SVM	0,41	0,36	0,44	0,37	0,37	0,32	0,76	0,70
Lasy losowe	0,40	0,36	0,43	0,36	0,36	0,31	0,79	0,71
Bayesowska	0,25	0,16	0,19	0,18	0,21	0,15	0,49	0,57
Liniowa (2)	0,26	0,16	0,18	0,18	0,23	0,16	0,45	0,49

Źródło: opracowanie własne.

**Tabela 19.** Zestawienie wartości błędu RMSPE – z uwzględnieniem nieruchomości podobnych

	Jelenia Góra – całość	Jelenia Góra – klasa 1	Jelenia Góra – klasa 2	Jelenia Góra – klasa 3	Warszawa – całość	Warszawa – klasa 1	Warszawa – klasa 2	Warszawa – klasa 3
Linowa	0,59	0,18	0,44	0,15	0,73	0,34	0,26	0,44
Lasso	0,53	0,18	0,43	0,15	0,64	0,31	0,24	0,38
SVM	0,56	0,17	0,49	0,16	0,66	0,35	0,27	0,36
Lasy losowe	0,56	0,17	0,48	0,17	0,69	0,34	0,28	0,42
Bayesowska	0,36	0,13	0,36	0,14	0,26	0,19	0,18	0,26
Linowa (2)	0,38	0,11	0,34	0,14	0,29	0,21	0,18	0,24

Źródło: opracowanie własne.

Tabela 20. Zestawienie wartości błędu MAPE – z uwzględnieniem nieruchomości podobnych

	Jelenia Góra – całość	Jelenia Góra – klasa 1	Jelenia Góra – klasa 2	Jelenia Góra – klasa 3	Warszawa – całość	Warszawa – klasa 1	Warszawa – klasa 2	Warszawa – klasa 3
Linowa	0,42	0,14	0,28	0,13	0,48	0,25	0,21	0,32
Lasso	0,38	0,13	0,26	0,13	0,43	0,23	0,20	0,29
SVM	0,41	0,13	0,29	0,14	0,44	0,27	0,21	0,27
Lasy losowe	0,40	0,13	0,28	0,14	0,43	0,25	0,22	0,30
Bayesowska	0,25	0,10	0,23	0,12	0,19	0,15	0,14	0,20
Linowa (2)	0,26	0,09	0,23	0,12	0,18	0,15	0,14	0,19

Źródło: opracowanie własne.

3.4 Struktura projektowanego systemu automatycznej wyceny nieruchomości

Proponowana struktura systemu automatycznej wyceny nieruchomości obejmuje trzy moduły, z których jeden moduł zawiera w sobie dalsze trzy podmoduły:

1. Moduł wejściowy
2. Moduł operacyjny, w ramach którego są:
 - 2.1. Podmoduł wyceny
 - 2.2. Podmoduł analityczno-prognostyczny
 - 2.3. Podmoduł archiwizujący
3. Moduł wyjściowy

Oto szczegółowy opis tych części systemu.

Moduł wejściowy

W module tym dane wejściowe wprowadzane są przez końcowego użytkownika systemu. Są to dane dotyczące wycenianej nieruchomości lub wycenianego portfela nieruchomości. Użytkownik będzie otrzymywał szeroką listę zmiennych i będzie mógł wprowadzić dane dotyczące wszystkich lub części tych zmiennych. Oprócz tego będzie możliwość wprowadzenia danych dotyczących dużej liczby nieruchomości w postaci pliku (np. csv lub xls).

Ponadto w module wejściowym użytkownik podaje wartości wyjściowe, którymi jest zainteresowany. Te wartości są wyznaczane w systemie. W pierwszej



wersji funkcjonowania systemu podstawowa wartość wyjściowa jest to wartość nieruchomości na dzień wyceny. W kolejnych wersjach, po uruchomieniu komponentu analityczno-prognostycznego, mogą to być inne wartości, w szczególności:

- ▶ prognoza wartości nieruchomości na określony dzień w przyszłości;
- ▶ prawdopodobny przedział wartości nieruchomości na określony dzień w przyszłości;
- ▶ oszacowanie wartości w przypadku zrealizowania się określonego scenariusza na rynku nieruchomości;
- ▶ oszacowanie przeszłej i przyszłej zmienności wartości nieruchomości.

Bardzo istotna jest struktura interfejsu wejściowego do komunikacji z końcowym użytkownikiem. Musi być dostosowana do niego. Planuje się, że będzie się ona różnić, inna będzie dla banku, inna dla inwestora, inna dla osoby indywidualnej.

Moduł wyjściowy

W module tym przedstawiane są wyniki wyjściowe systemu. Otrzymywane są wartości tych zmiennych, którymi jest zainteresowany użytkownik, a które zostały określone przez niego w module wejściowym. W pierwszej wersji funkcjonowania systemu jest to wartość nieruchomości na dzień wyceny. Jest oczywiście możliwość otrzymania wyników w pliku (np. csv lub xls).

Bardzo istotna jest struktura interfejsu wyjściowego do komunikacji z końcowym użytkownikiem. Musi być dostosowana do niego. Planuje się, że będzie się ona różnić, inna będzie dla banku, inna dla inwestora, inna dla osoby indywidualnej.

Moduł operacyjny

Docelowo ten moduł zawierać będzie trzy podmoduły:

- ▶ podmoduł wyceny;
- ▶ podmoduł analityczno-prognostyczny;
- ▶ podmoduł archiwizujący.

W pierwszej wersji będzie funkcjonować podmoduł wyceny, gdyż głównym celem systemu jest automatyczna wycena nieruchomości mieszkaniowych. Od początku będzie również funkcjonować moduł archiwizujący.

Podmoduł wyceny

Ten podmoduł ma na celu wycenę nieruchomości na dowolny dzień wyceny.

Cechy tej części systemu są następujące:

- ▶ System będzie połączony z bazą danych o nieruchomościach (AMRON), czyli w dowolnym momencie może pobierać dowolne dane z tej bazy.
- ▶ Na potrzeby działania systemu baza danych będzie podzielona na zbiory danych odpowiadających nieruchomościom podobnym. Pierwsze kryterium podziału to miasto, a zatem dla każdego miasta będzie osobny podzbiór danych. Drugie kryterium to obszar w ramach dużego miasta. Jako trzecie kryterium można zastosować zbiór cech różniących nieruchomości na danym obszarze (w danym mieście). Tutaj może być zastosowana metoda zaliczana do wskazanych wcześniej metod klasyfikacji, czyli grupy metod uczenia maszynowego.
- ▶ W systemie będzie oprogramowana pewna liczba metod wyceny spełniających warunek metod analizy statystycznej rynku (podejście porównawcze). Liczba ta jest większa niż liczba metod przyjętych do testowania. Oznacza to, że dla zbioru danych o dowolnej nieruchomości będzie możliwość wyznaczenia jej wartości.
- ▶ Podmoduł wyceny może działać w dwóch trybach: tryb zadany przez użytkownika oraz tryb domyślny.
- ▶ Tryb zadany przez użytkownika oznacza, że wycena dokonywana jest przy określeniu przez użytkownika przynajmniej jednej z dwóch charakterystyk: pierwsza to metoda wybrana przez niego, druga to zestaw zmiennych dotyczących nieruchomości.
- ▶ Tryb domyślny będzie zdecydowanie częściej stosowany. Przewiduje się, że w pierwszej wersji funkcjonowania systemu będzie to jedyny tryb. Tryb domyślny oznacza, że dla wyznaczonej nieruchomości – przy zadanych wartościach zmiennych ją charakteryzujących – wyznaczona będzie jej wartość.
- ▶ Na potrzeby wyceny wskazanej nieruchomości zastosowana zostanie jedna metoda spośród dostępnego zestawu. Metoda ta przy zastosowanym kryterium (np. minimalizacja jednego z podanych powyżej mierników – wzory (8)–(12)) jest w danym dniu optymalna.



- ▶ Co pewien czas (np. raz na miesiąc) system będzie dokonywał uaktualnienia podmodułu wyceny. To uaktualnienie przebiegać będzie w dwóch etapach:
- ▶ porównanie obecnie wyliczonych wartości z wartościami prognozowanymi w przeszłości;
- ▶ przechowywanie wyników odchyleń prognozytycznych.

Etap 1. Wyznaczenie klas nieruchomości podobnych.

Dla każdego obszaru wyróżnione zostaną klasy nieruchomości podobnych. Obszary (miasta, dzielnice miast itp.) powinny zostać określone metodą ekspercką. Podobna metoda powinna być zastosowana przy wyborze zmiennych na potrzeby klasyfikacji. Z kolei klasy nieruchomości podobnych dla danego obszaru i dla zadanego zestawu zmiennych zostaną wyznaczone za pomocą metod klasyfikacji.

Etap 2. Wyznaczenie modelu wyceny dla każdej klasy nieruchomości podobnych.

W tym etapie pod uwagę zostaną wzięte różne możliwe metody wyceny (np. N metod) i różne możliwe zestawy zmiennych (np. M zmiennych). W ten sposób otrzymujemy $K = N \times M$ modeli wyceny. System wybiera ten spośród K modeli, który jest w dniu uaktualnienia modułu operacyjnego (np. raz w miesiącu) optymalny.

Ten wybrany w momencie uaktualnienia systemu model wyceny jest stosowany w odniesieniu do wyceny nieruchomości przedstawianej w module wejścia systemu, aż do kolejnego uaktualnienia systemu.

9. Co pewien czas (np. raz na kwartał) system będzie dokonywał automatycznie uaktualnienia wartości nieruchomości zawartych w bazie danych. W tym celu zostanie wykorzystany aktualny model wyceny dla danego zbioru nieruchomości podobnych, który to model zostaje uzyskany w procesie aktualizacji podmodułu wyceny. Oczywiście uaktualnienie wartości nieruchomości zawartych w bazie danych może mieć miejsce w momencie uaktualnienia podmodułu wyceny (czyli np. raz w miesiącu).

Podmoduł archiwizujący

Ten podmoduł będzie miał charakter techniczny, gdyż gromadził będzie historię operacji w systemie. Umożliwi to dostęp do wartości nieruchomości w dowolnym dniu w przeszłości (od momentu uruchomienia systemu). Pozwoli to m.in. na:

Podmoduł analityczno-prognostyczny

Ten podmoduł jest naturalnym elementem systemu, który powinien być uruchomiony po pomyślnej weryfikacji działania pierwszej wersji systemu.

W podmodule tym będą wykonywane następujące zadania:

1. Wyznaczenie dla dowolnej wycenianej nieruchomości mieszkaniowej przedziału wyceny, tzn. wskazanie dolnego końca przedziału wartości i górnego końca przedziału wartości, który to przedział z dużym prawdopodobieństwem zawierać będzie wartość nieruchomości; pozwoli to na ocenę ryzyka modelu wyceny.
2. Wyznaczenie prognozy wartości nieruchomości – tutaj na etapie tworzenia tego podmodułu będzie określone, jaki jest maksymalny możliwy do przyjęcia horyzont prognozy. Przy tym prognoza może być tworzona:
 - a) na podstawie trendów;
 - b) na podstawie informacji fundamentalnych, zwłaszcza makroekonomicznych (w tym przypadku należy oprócz bazy AMRON skorzystać z bazy zawierającej dane fundamentalne);
 - a) z uwzględnieniem obu powyższych metod.
3. Wyznaczenie prognozy przeciętnej wartości nieruchomości na dowolnym lokalnym rynku.
4. Wyznaczenie parametru zmienności cen dowolnej nieruchomości, co pozwoli na ocenę ryzyka ceny nieruchomości, jak również na ocenę zmian w czasie jej wartości.
5. Wyznaczenie prognozy wartości zabezpieczenia dla portfela kredytów.
6. Analiza *ex post* zmian wartości pojedynczej nieruchomości na podstawie danych historycznych zarchiwizowanych w systemie.



7. Analiza *ex post* zmian wartości portfela zabezpieczeń na podstawie danych historycznych zachwizowanych w systemie.

Oczywiście możliwe jest rozszerzenie zbioru zadań wykonywanych przez podmoduł analityczno-prognostyczny.

3.5 Funkcjonalność systemu dla różnych interesariuszy

Funkcjonalność systemu – odbiorca: bank

Bank będzie mógł ko Δ zystać z następujących elementów funkcjonalności systemu:

1. Informacje dotycz $\square$ ce pojedynczego kredytu:

- $\boxtimes$ wstępna wycena nieruchomości w momencie wydawania formularza informacyjnego;
- $\boxtimes$ wyce $\Leftrightarrow$ a nieruchomości w momencie podpisywania umowy kredytu (weryfikacja dostarczonego operat $\blacktriangleright$ szacunkowego);
- $\boxtimes$ weryfikacja wartości nieruchomości określonej przez rzeczoznawcę w dowolnym momencie;
- $\boxtimes$ wycena nieruchomości jako przedmiotu zabezpieczenia w sytuacji wystąpienia opóźnień w obsłudze kredytu;
- $\boxtimes$ ocena poziomu zabezpieczenia ekspozycji kredytowej.

2. Informacje dotyczące portfela kredytów:

- $\blacktriangleright$ wartość portfela zabezpieczeń kredytów w dowolnym momencie wyceny, jak również automatyczna informacja o tej wartości w ustalonych odstępach czasowych, np. raz w miesiącu;
- $\blacktriangleright$ wycena *cover pool* w sytuacji transferu wierzytelności z banku uniwersalnego do banku hipotecznego (emitenta listów zastawnych);
- $\blacktriangleright$ ocena poziomu zabezpieczenia ekspozycji kredytowej portfela.

Po rozszerzeniu systemu o podmoduł analityczno-prognostyczny bank będzie mógł korzystać z następujących dodatkowych elementów funkcjonalności systemu:

- $\blacktriangleright$ wyznaczenie prognozy wartości pojedynczej nieruchomości;
- $\blacktriangleright$ wyznaczenie przyszłej wartości pojedynczej nieruchomości przy założeniu scenariusza kształtowania się sytuacji rynkowej – wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych (tzw. *stress*);
- $\blacktriangleright$ wyznaczenie prognozy wartości portfela zabezpieczeń;
- $\blacktriangleright$ wyznaczenie przyszłej wartości portfela zabezpieczeń przy założeniu scenariusza kształtowania się sytuacji rynkowej – wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych (tzw. *stress*).

Funkcjonalność systemu – odbiorca: fundusz inwestycyjny

Fundusz inwestycyjny będzie mógł korzystać z następujących elementów funkcjonalności systemu.

1. Informacje dotyczące pojedynczej nieruchomości:
 - $\blacktriangleright$ wycena na dany dzień;
 - $\blacktriangleright$ przeszłe wartości nieruchomości.
2. Informacje dotyczące portfela nieruchomości:
 - $\blacktriangleright$ są to te same informacje, co w przypadku pojedynczej nieruchomości.

Po rozszerzeniu systemu o podmoduł analityczno-prognostyczny fundusz inwestycyjny będzie mógł korzystać z następujących dodatkowych elementów funkcjonalności systemu:

- $\blacktriangleright$ wyznaczenie prognozy wartości pojedynczej nieruchomości;
- $\blacktriangleright$ wyznaczenie przyszłej wartości pojedynczej nieruchomości przy założeniu scenariusza kształtowania się sytuacji rynkowej – wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych (tzw. *stress*);
- $\blacktriangleright$ wyznaczenie prognozy wartości portfela nieruchomości;



- ▶ wyznaczenie przyszłej wartości portfela nieruchomości przy założeniu scenariusza kształtowania się sytuacji rynkowej – wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych (tzw. *stress*).

Funkcjonalność systemu – odbiorca: zakład ubezpieczeń

Zakład ubezpieczeń, który sprzedaje ubezpieczenia kredytów mieszkaniowych, będzie mógł korzystać z następujących elementów funkcjonalności systemu.

1. Informacje dotyczące pojedynczej nieruchomości:
 - ▶ wycena na dany dzień.
2. Informacje dotyczące portfela nieruchomości:
 - ▶ są to te same informacje, co w przypadku pojedynczej nieruchomości.

Po rozszerzeniu systemu o podmoduł analityczno-prognostyczny zakład ubezpieczeń będzie mógł korzystać z następujących dodatkowych elementów funkcjonalności systemu:

- ▶ wyznaczenie prognozy wartości pojedynczej nieruchomości;
- ▶ wyznaczenie przyszłej wartości pojedynczej nieruchomości przy założeniu scenariusza kształtowania się sytuacji rynkowej – wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych (tzw. *stress*);
- ▶ wyznaczenie prognozy wartości portfela nieruchomości;
- ▶ wyznaczenie przyszłej wartości portfela nieruchomości przy założeniu scenariusza kształtowania się sytuacji rynkowej – wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych (tzw. *stress*).

Funkcjonalność systemu – odbiorca: gospodarstwo domowe

Gospodarstwo domowe będzie mogło skorzystać z następujących elementów funkcjonalności systemu:

- ▶ wstępna wycena nieruchomości;

- ▶ weryfikacja wartości nieruchomości określonej przez rzeczoznawcę;
- ▶ wycena nieruchomości w dowolnym momencie.

Po rozszerzeniu systemu o podmoduł analityczno-prognostyczny gospodarstwo domowe będzie mogło korzystać z następujących dodatkowych elementów funkcjonalności systemu:

- ▶ wyznaczenie prognozy wartości pojedynczej nieruchomości;
- ▶ wyznaczenie przyszłej wartości pojedynczej nieruchomości przy założeniu scenariusza kształtowania się sytuacji rynkowej (wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych – tzw. *stress*);
- ▶ wybór z zestawu mieszkań o podanych cechach mieszkania o najniższej wartości.

Funkcjonalność systemu – odbiorca: deweloper

Deweloper będzie mógł skorzystać z następujących elementów funkcjonalności systemu:

- ▶ wycena nieruchomości;
- ▶ selekcja nieruchomości na danym obszarze i ich wycena.

Po rozszerzeniu systemu o podmoduł analityczno-prognostyczny deweloper będzie mógł korzystać z następujących dodatkowych elementów funkcjonalności systemu:

- ▶ wyznaczenie prognozy wartości nieruchomości na danym obszarze;
- ▶ wyznaczenie przyszłej wartości pojedynczej nieruchomości przy założeniu scenariusza kształtowania się sytuacji rynkowej (wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych – tzw. *stress*).

Funkcjonalność systemu – inni odbiorcy

Jak wskazywaliśmy, innymi odbiorcami systemu mogą być:

- ▶ decydenci gospodarczy (na poziomie centralnym i lokalnym), prowadzący politykę mieszkaniową;

- ▶ decydenci gospodarczy, analizujący koniunkturę gospodarczą;
- ▶ instytucje badawcze, opracowujące raporty o rynku nieruchomości;
- ▶ instytucje naukowe, prowadzące badania naukowe dotyczące nieruchomości.

Będą oni mogli korzystać z następujących elementów funkcjonalności systemu.

1. Informacje dotyczące pojedynczej nieruchomości:

- ▶ wycena na dany dzień;
- ▶ przeszłe wartości nieruchomości.

2. Informacje dotyczące portfela nieruchomości:

- ▶ są to te same informacje, co w przypadku pojedynczej nieruchomości.

Po rozszerzeniu systemu o podmoduł analityczno-prognostyczny będą mogli korzystać z następujących dodatkowych elementów funkcjonalności systemu:

- ▶ wyznaczenie prognozy wartości pojedynczej nieruchomości;
- ▶ wyznaczenie przyszłej wartości pojedynczej nieruchomości przy założeniu scenariusza kształtowania się sytuacji rynkowej – wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych (tzw. *stress*);
- ▶ wyznaczenie prognozy wartości portfela nieruchomości;
- ▶ wyznaczenie przyszłej wartości portfela nieruchomości przy założeniu scenariusza kształtowania się sytuacji rynkowej – wstępnie przewiduje się kilka bazowych scenariuszy, w tym scenariusz warunków skrajnych (tzw. *stress*).

Należy dodać jeszcze jedną cechę zwiększającą funkcjonalność systemu. Chodzi tu o możliwość zastosowania systemu do masowej wyceny nieruchomości, co może mieć znaczenie w przypadku wprowadzenia powszechnego podatku od wartości (*ad valorem*) nieruchomości.

Funkcjonalność systemu – interface

Funkcjonalność systemu zależy również od interface, czyli sposobu komunikowania się końcowego użytkownika z systemem. W momencie tworzenia systemu należy zasięgnąć informacji od końcowych użytkowników, przede wszystkim od banków, o rodzajach informacji, którymi użytkownik jest zainteresowany. Chodzi tu o to, aby te informacje, które generuje system, były dostępne dla użytkownika bez konieczności dodatkowego ich przetwarzania przez niego.

3.6 Spełnianie wymogów regulacyjnych przez system

System będzie spełniał wymogi regulacyjne.

Po pierwsze, wycena jest dokonywana metodami analizy statystycznej rynku. Oznacza to zgodność z dwoma podstawowymi aktami prawnymi, którymi są:

- ▶ Ustawa z dnia 21 sierpnia 1997 roku o gospodarce nieruchomościami;
- ▶ Rozporządzenie Rady Ministrów z dnia 21 września 2004 roku w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego.

Po drugie, spełnione są wymogi nadzorcze zawarte w przedstawionej przez Komisję Nadzoru Finansowego Rekomendacji J dotyczącej zasad gromadzenia i przetwarzania przez banki danych o nieruchomościach.

Oto odniesienie – w postaci komentarzy – do szczegółowych rekomendacji zawartych w ogólnej Rekomendacji J.

I. Gromadzenie danych o rynku nieruchomości

Rekomendacja 1

Niepoprawnie lub mało ostrożnie przyjmowane przez banki wartości zabezpieczeń w zakresie istotnego co do wartości i udziału portfela ekspozycji kredytowych zabezpieczonych hipotecznie może powodować występowanie ryzyka systemowego. Dlatego też niezbędnym elementem polityki zarządzania ryzykiem w banku, umożliwiającym ograniczanie ryzyka systemowego, jest zarządzanie ryzykiem związanym z nie-



ruchomościami przyjmowanymi jako zabezpieczenie hipoteczne. Powinno to się odbywać na podstawie danych dotyczących rynku nieruchomości gromadzonych w wewnętrznych (własnych) lub zewnętrznych (międzybankowych) bazach danych.

Komentarz: często aktualizowana wycena nieruchomości uzyskiwana na podstawie systemu wykorzystującego dane z bazy AMRON pozwoli na dostosowanie do tej rekomendacji.

Rekomendacja 2

Gromadzone przez banki dane powinny w wystarczającym stopniu pozwolić na ocenę i monitorowanie wartości zabezpieczenia na nieruchomości, prowadzenie analiz rynku nieruchomości, a w przypadku banków istotnie zaangażowanych na ocenę ryzyka zmiany wartości zabezpieczeń na nieruchomościach.

Komentarz: projektowany system automatycznej wyceny nieruchomości uzupełniony o podmoduł analityczno-prognostyczny pozwoli na dostosowanie do tej rekomendacji.

Rekomendacja 3

Istotne zaangażowanie banków w finansowanie nieruchomości wiąże się z koniecznością zapewnienia długoterminowego finansowania dla tego obszaru działalności. Tworzenie oraz aktywne korzystanie przez banki z baz danych gromadzących informacje na temat rynku nieruchomości powinno umożliwiać rozwój systemu pozyskiwania długoterminowego finansowania przez banki z wykorzystaniem instrumentów dłużnych w celu budowy pozycji płynnościowej banków. Banki powinny szczegółowo analizować strukturę terminową źródeł finansowania i dostosowywać ją do struktury terminowej aktywów, dywersyfikując źródła finansowania działalności w celu ograniczania ryzyka płynności.

Komentarz: dostosowanie do tej rekomendacji dla banków jest łatwiejsze przez aktualizowanie wyceny tej części aktywów, która jest zgromadzona w kredytach mieszkaniowych.

Rekomendacja 4

Zasady prowadzenia i korzystania z baz danych o rynku nieruchomości powinny zostać określone w polityce zarządzania ryzykiem kredytowym oraz opisane w procedurach zatwierdzonych zgodnie z obowiązującymi

w banku zasadami wdrażania przepisów wewnętrznych w ramach jednolitego systemu zarządzania ryzykiem.

Komentarz: ta rekomendacja nie dotyczy bezpośrednio systemu wyceny, aczkolwiek wykorzystanie projektowanego systemu automatycznej wyceny nieruchomości przez banki będzie miało bardzo pozytywny wpływ na system zarządzania ryzykiem w banku.

II. Tworzenie baz danych o rynku nieruchomości

Rekomendacja 5

Bazy danych o rynku nieruchomości powinny zawierać dane pozwalające na klasyfikowanie nieruchomości do grup o wysokim poziomie podobieństwa cech wpływających na ich wartość.

Komentarz: prowadzona w systemie aktualizacja wyceny (np. co miesiąc), w której pierwszym etapem jest uaktualnienie klasyfikacji nieruchomości mieszkaniowych na grupy nieruchomości podobnych, jest dopasowana do wymogów tej rekomendacji.

Rekomendacja 6

Szczególnie istotne jest, aby banki aktywnie włączyły się w system zewnętrznej (międzybankowej) wymiany informacji w zakresie rynku nieruchomości, tworząc w tym celu wspólną dla wszystkich banków zewnętrzną (międzybankową) bazę danych, zasilaną również z wiarygodnych źródeł pozabankowych. Baza ta powinna być systematycznie zasilana przez wszystkie banki objęte niniejszą rekomendacją.

Komentarz: wykorzystanie przez system bazy AMRON, która spełnia warunki tej rekomendacji oraz możliwość wykorzystania danych fundamentalnych w podmodule analityczno-prognostycznym pozwoli na dostosowanie do wymogów tej rekomendacji.

III. Wiarygodność baz danych o rynku nieruchomości

Rekomendacja 7

Wykorzystywaną przez bank bazę danych o rynku nieruchomości uznaje się za wiarygodną, jeżeli:



- ▶ jest na bieżąco zasilana informacjami wyłącznie ze źródeł uznanych za wiarygodne, zwłaszcza określonych w rekomendacji 1, 2, 3.
- ▶ poszczególne informacje gromadzone są w sposób wystandaryzowany,
- ▶ jest zagwarantowana wysoka jakość informacji zwrotnych generowanych w ramach bieżącego wykorzystania bazy danych,
- ▶ posiada odpowiednio duży zasób informacji (historycznych i aktualnych) o poszczególnych, lokalnych rynkach nieruchomości,
- ▶ system informatyczny obsługujący bazę danych charakteryzuje się odpowiednim poziomem wydajności i bezpieczeństwa, zgodnie z dobrą i uznaną praktyką w tym zakresie,
- ▶ zgromadzone informacje pozwalają na dokonywanie oceny i monitorowania wartości zabezpieczenia na nieruchomości oraz prowadzenie analiz i generowanie raportów dotyczących zarówno rynków lokalnych, jak i całego rynku nieruchomości,
- ▶ jest obsługiwana przez osoby/komórki organizacyjne o odpowiednich kompetencjach,
- ▶ bank posiada i stosuje odpowiednie procedury, zatwierdzone zgodnie z obowiązującymi w banku zasadami wdrażania przepisów wewnętrznych, regulujące wykorzystywanie baz danych oraz uznawanie ich za wiarygodne, zarówno w odniesieniu do baz wewnętrznych, jak i zewnętrznych – jeżeli są przez bank wykorzystywane.

Komentarz: baza danych AMRON, na podstawie której będzie działał projektowany system automatycznej wyceny nieruchomości, spełnia warunki tej rekomendacji. Dodatkowo sprzyjać temu będzie aktualizacja wycen nieruchomości z bazy AMRON dokonywana przez system automatycznej wyceny nieruchomości.

Rekomendacja 8

Bank powinien raz do roku weryfikować wiarygodność wykorzystywanych baz danych o rynku nieruchomości. Sposób weryfikacji, kryteria i osoby/komórki organizacyjne banku upoważnione do jej przeprowadzania powinny być wskazane w odpowiednich procedurach, zatwierdzonych zgodnie

z obowiązującymi w banku zasadami wdrażania przepisów wewnętrznych.

Komentarz: działanie systemu, w którym często (częściej niż raz w roku) będzie dokonywana aktualizacja wycen nieruchomości zawartych w bazie AMRON, pozwoli na dostosowanie do wymogów tej rekomendacji.

IV. Korzystanie z baz danych o rynku nieruchomości

Rekomendacja 9

Baza danych o rynku nieruchomości powinna pozwolić bankowi na ocenę i monitorowanie wartości zabezpieczenia na nieruchomości dla większości analizowanych przez bank wniosków.

Komentarz: wyniki wyjściowe systemu będą zawierać wartości zabezpieczeń kredytów mieszkaniowych, co oznacza bezpośrednie dostosowanie do wymogów tej rekomendacji.

Rekomendacja 10

Bank powinien posiadać procedury oceny i monitorowania wartości zabezpieczenia na nieruchomości, dla których nie jest możliwe wykonanie ocen na podstawie baz danych o rynku nieruchomości tworzonych zgodnie z niniejszą rekomendacją.

Komentarz: zakłada się, że bazy danych wykorzystywanych przez banki i system automatycznej wyceny nieruchomości mieszkaniowych będą tworzone w zgodzie z Rekomendacją J, a zatem dla wszystkich nieruchomości mieszkaniowych będzie to możliwe za pośrednictwem projektowanego systemu.

Rekomendacja 11

W przypadku banku istotnie zaangażowanego, dla właściwej oceny ryzyka zmiany wartości zabezpieczeń na nieruchomościach, niezbędne jest stosowanie modeli statystycznych opartych na danych pochodzących z wiarygodnych zewnętrznych (międzybankowych) lub wewnętrznych baz danych.

Komentarz: w tym zakresie spełnienie wymogów rekomendacji jest ewidentne, gdyż wykorzystywane będą metody analizy statystycznej rynku oparte na danych bazy AMRON.

Rekomendacja 12

Zapewnienie odpowiedniej jakości modeli oceny ryzyka zmiany wartości zabezpieczeń na nieruchomościach stosowanych przez bank wymaga, aby ich wykorzystanie i skuteczność podlegały regularnej weryfikacji w ramach monitoringu lub walidacji.

Komentarz: weryfikacja modeli oceny ryzyka przez banki będzie łatwiejsza z uwagi na okresową aktualizację wycen zabezpieczeń dokonywaną przez system automatycznej wyceny nieruchomości.

Ponadto należy wspomnieć o innych zaletach wykorzystywanej przez system bazy AMRON. Są nimi:

- ▶ pozytywne doświadczenia z kilkunastoletniej historii tej bazy;
- ▶ jednolita struktura danych w tej bazie;
- ▶ szeroki zasięg tej bazy;
- ▶ relatywnie niższe koszty użytkowania tej bazy.

Jest jeszcze jeden argument, który nie wynika z polskich regulacji prawnych, lecz ze standardów światowych. Jak wcześniej wspomniano, w zakresie systemów automatycznej wyceny nieruchomości

w Stanach Zjednoczonych (gdzie te systemy są najszerszej stosowane) opublikowano standard jakości dla tych systemów. Jest to „Standard on Automated Valuation Models (AVMs)”. W dokumencie tym bardzo szczegółowo opisano pożądane cechy modeli (systemów) automatycznej wyceny nieruchomości. Na ogólnym poziomie sformułowano pięć takich cech:

- ▶ szerokie zastosowanie;
- ▶ wykorzystanie dostatecznej informacji statystycznej;
- ▶ przejrzystość;
- ▶ zapewnienie jakości;
- ▶ zaufanie użytkowników.

Przedstawiony wcześniej opis systemu pozwala na stwierdzenie o spełnianiu przez ten system cechy pierwszej i drugiej. Cecha trzecia powinna być spełniona przez dostosowanie właściwego interfejsu w systemie. Cecha czwarta powinna być spełniona przez weryfikację działania systemu w początkowym okresie jego funkcjonowania. To zaś automatycznie będzie oznaczać spełnienie cechy piątej.



Dalsze proponowane działania





W przypadku akceptacji koncepcji systemu zachodzi konieczność rozwijania go. Zespół jest gotowy do pomocy w tych pracach.

Jest to możliwe co najmniej w następujących etapach:

1. Etap konstrukcji podmodułu wyceny:

- ▶ oferta zestawu możliwych metod automatycznej wyceny spośród metod uczenia maszynowego;
- ▶ opracowanie metody klasyfikacji na zbiory nieruchomości podobnych;
- ▶ nadzór nad implementacją i testowaniem metod.

2. Etap konstrukcji podmodułu analityczno-prognostycznego w zakresie co najmniej następujących prac:

- ▶ opracowanie metody konstrukcji przedziału wyceny;

- ▶ opracowanie metody prognozy wartości nieruchomości:

a) na podstawie trendów;

b) na podstawie informacji fundamentalnych, zwłaszcza makroekonomicznych;

c) z uwzględnieniem obu metod.

- ▶ opracowanie metody prognozy przeciętnej wartości nieruchomości na dowolnym lokalnym rynku;

- ▶ opracowanie metody wyznaczania parametru zmienności cen dowolnej nieruchomości;

- ▶ opracowanie metod analizy statystycznej danych historycznych dotyczących nieruchomości;

- ▶ opracowanie metody prognozowania wartości zabezpieczenia dla portfela kredytów;

- ▶ opracowanie indeksu cen nieruchomości dla grup nieruchomości podobnych.



Literatura





- Cymerman R., Hopfer A. (2008), *System, zasady i procedury wyceny nieruchomości*, Polska Federacja Stowarzyszeń Rzeczoznawców Majątkowych, Warszawa.
- Czaja J., (1999), *Metody i systemy określania wartości nieruchomości*, Uczelniane Wydawnictwa Naukowo-Dydaktyczne, Kraków.
- Czaja J., Parzych P., (2008), *Szacowanie rynkowej wartości nieruchomości w aspekcie międzynarodowych standardów wyceny*, Wydawnictwo Stowarzyszenia Naukowego im. St. Staszica, Kraków.
- Downie, M.L., Robson, G. (2008), *Automated Valuation Models: An International Perspective*, Council of Mortgage Lenders, London.
- Dydenko J. (2006) (red.), *Szacowanie nieruchomości*, Dom Wydawniczy ABC, Gdynia.
- Everitt B., Landau S., Leese M., Stahl D. (2011), *Cluster Analysis*, Wiley, New York.
- Foryś I., Kokot S., (2008), *Preferencje potencjalnych nabywców w określaniu wag cech rynkowych nieruchomości*, Studia i Materiały Towarzystwa Naukowego Nieruchomości, 16, Olsztyn.
- Frew J., Jud G.D. (2003), *Estimating the value of Apartment Buildings*, Journal of Real Estate Research, vol. 25 No. 1.
- Haas, G. C. (1922), *Sale prices as a basis for farm land appraisal*, Technical Bulletin 9, St. Paul: The University of Minnesota Agricultural Experiment Station.
- Hopfer A., Jędrzejewski H., Żróbek R., Żróbek S. (2001), *Podstawy wyceny nieruchomości*, Twigger, Warszawa.
- Hopfer A., Jędrzejewski H., Żróbek S., Żróbek R. (1994), *Wycena nieruchomości i przedsiębiorstw*. Tom 1. *Szacowanie nieruchomości*, Twigger, Warszawa.
- Hozier J. (2001) (red.), *Nieruchomości, przedsiębiorstwa, wycena, analizy*. Tom I, Uniwersytet Szczeciński, Szczecin.
- Prystupa M., Mączyńska E., Rygiel K. (2009), *Ile jest warta nieruchomość*, Poltext, Warszawa.
- Rattermann M.R. (2007), *Valuation by Comparison. Residential Analysis and Logic*, Appraisal Institute, Chicago.
- Rattermann M.R. (2008), *Study Guide to Appraising Residential Properties*, Appraisal Institute, Chicago.
- Rozporządzenie Rady Ministrów z dnia 21 września 2004 roku w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego (Dz.U. Nr 207, poz. 2109 ze zm.).
- Standard on Automated Valuation Models (AVMs) (2018), International Association of Assessing Officers, Kansas City.
- Ustawa z dnia 21 sierpnia 1997 roku o gospodarce nieruchomościami (Dz.U. z 2018 r., poz. 2204 ze zm.).
- Żróbek S. (2002) (red.), *Określanie wartości rynkowej nieruchomości*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn.



Raport opracowany na zlecenie Programu Analityczno-Badawczego
Fundacji Warszawski Instytut Bankowości

