

UNIVERSITATEA BABEȘ-BOLYAI CLUJ-NAPOCA

FACULTATEA DE ȘTIINȚE ECONOMICE ȘI GESTIUNEA AFACERILOR

Școala doctorală de Științe Economice și Gestiunea Afacerilor

REZUMATUL TEZEI DE DOCTORAT CU TITLUL

***INTEGRAREA MODELELOR DE ÎNVĂȚARE AUTOMATĂ ÎN EVALUAREA
PROPRIETĂȚILOR IMOBILIARE***

Coordonator științific:

Prof. univ. dr. habil. Adela DEACONU

Doctorand:

Silviu-Ionuț BĂBȚAN

Cluj-Napoca, 2025

CUPRINSUL TEZEI DE DOCTORAT

LISTA ABREVIERILOR _____

LISTA FIGURILOR _____

LISTA GRAFICELOR _____

LISTA TABELELOR _____

MULȚUMIRI _____

INTRODUCERE _____

CAPITOLUL 1 – EVALUAREA ACTIVELOR ÎN DIFERITE SCOPURI ERROR! BOOKMARK NOT DEFINED.

1.1 DELIMITĂRI – GENERALITĂȚI _____

1.1.1 Scurt istoric al evaluării _____

1.1.2 Conceptele de „preț”, „cost” și „valoare” _____

1.1.3 Obiectul evaluării _____

1.2 ABORDĂRI, METODE ȘI TEHNICI DE EVALUARE _____

1.2.1 Abordarea prin piață _____

1.2.1.1 Elemente determinante în analiza și variația prețurilor imobiliare _____

1.2.1.2 Identificarea și cuantificarea corecțiilor _____

1.2.1.3 Comparație cu Standardele de Evaluare a bunurilor aplicabile în România (SEV) _____

1.2.2 Abordarea prin venit _____

1.2.2.1 Capitalizarea directă și metoda actualizării fluxurilor de numerar _____

1.2.3 Abordarea prin cost _____

1.2.3.1 Procedura abordării prin cost _____

1.2.3.2 Tipuri de depreciere _____

1.3 SCOPUL ȘI UTILITATEA EVALUĂRII _____

1.4 CONCLUZIILE CAPITOLULUI 1 _____

CAPITOLUL 2 – ANALIZA LITERATURII PRIVIND EVALUĂRILE AUTOMATE _____

2.1 DESCRIEREA MODULUI DE ANALIZĂ A LITERATURII _____

2.2 ANALIZA BIBLIOMETRICĂ A LITERATURII _____

2.3 ANALIZA SISTEMATICĂ A LITERATURII _____

2.3.1 Conținutul și clasificările specifice AVM _____

2.3.2 Integrarea AVM în procesul de evaluare a proprietăților imobiliare _____

2.3.3 Cadrul metodologic al AVM _____

2.3.4 Beneficii și limitări în utilizarea AVM _____

2.4 CONCLUZIILE CAPITOLULUI 2 _____

CAPITOLUL 3 – MODELE DE ÎNVĂȚARE AUTOMATĂ ȘI APLICABILITATEA ACESTORA ÎN
EVALUAREA PROPRIETĂȚILOR IMOBILIARE _____

3.1 ÎNVĂȚAREA AUTOMATĂ ȘI APLICABILITATEA EI ÎN EVALUAREA PROPRIETĂȚILOR IMOBILIARE _____

3.2 MODELE DE ÎNVĂȚARE AUTOMATĂ ÎN EVALUAREA PROPRIETĂȚILOR IMOBILIARE _____

3.3.1 Regresia liniară _____

3.3.2 Arborele de decizie _____

3.3.3 Pădurea aleatorie (Random Forest) _____

3.3.4 Rețele neuronale artificiale (Artificial Neuronal Network – ANN) _____

3.3.5 Extreme Gradient Boosting (XGBoost) _____

3.4 EVALUAREA PERFORMANȚEI MODELELOR DE ÎNVĂȚARE AUTOMATĂ _____

3.4.1	Abordarea indicatorului-cheie Eroare Medie Pătratică (Root Mean Squared Error – RMSE)	___
3.4.2	Abordarea indicatorului-cheie Eroarea Medie Absolută (Mean Absolute Error – MAE)	_____
3.4.3	Abordarea indicatorului-cheie aferent Coeficientul de Determinare (R^2)	_____
3.4.4	Abordarea indicatorilor-cheie F-statistic și P-value	_____
3.5	MODUL DE LUCRU ÎN PROGRAMUL INFORMATIC RSTUDIO	_____
3.6	CONCLUZIILE CAPITOLULUI 3	_____
CAPITOLUL 4 – EVALUAREA PERFORMANTELOR MODELELOR DE ÎNVĂȚARE AUTOMATĂ ÎN PREDICȚIA PREȚURILOR PROPRIETĂȚILOR IMOBILIARE DE TIP REZIDENȚIAL		
4.1	MOTIVAȚIA ȘI ETAPELE CONSTRUCȚIEI MODELELOR	_____
4.2	SETUL DE DATE	_____
4.3	VARIABILELE	_____
4.4	CONSTRUIREA MODELELOR SI VALIDAREA LOR EMPIRICĂ	_____
4.4.1	Regresia liniară	_____
4.4.2	Arborii de decizie	_____
4.4.3	Păduri aleatorii	_____
4.4.4	Rețele Neuronale Artificiale	_____
4.4.5	Extreme Gradient Boosting	_____
4.5	REZULTATE OBTINUTE	_____
4.5.1	Analiza comparativă a metodelor aplicate	_____
4.5.2	Performanța modelului XGBoost în estimarea prețurilor proprietăților imobiliare	_____
4.5.3	Evaluarea finală a modelului XGBoost pe setul de testare	_____
4.6.1	Discuții, interpretări și limitări ale studiului de caz	_____
4.6.2	Concluzii finale	_____
CONCLUZII ȘI DIRECȚII VIITOARE DE CERCETARE		
BIBLIOGRAFIE		

Introducere

Lucrarea abordează utilizarea metodelor de învățare automată în evaluarea proprietăților imobiliare, având ca scop îmbunătățirea procesului tradițional de evaluare, care este adesea costisitor, lent și influențat de factori subiectivi. În contextul unei piețe imobiliare dinamice și digitalizate, introducerea modelelor automate de evaluare (AVM) devine o necesitate practică și strategică.

Se evidențiază faptul că AVM-urile, prin integrarea inteligenței artificiale, pot oferi rezultate rapide, obiective și consistente, contribuind la reducerea costurilor și eliminarea erorilor umane. Lucrarea propune aplicarea și compararea a cinci metode de predicție – regresia liniară, arborii de decizie, pădurile aleatorii, rețelele neuronale și XGBoost – pe un eșantion de apartamente din Cluj-Napoca.

Cercetarea este motivată de nevoia de soluții sustenabile și inovatoare în evaluarea imobiliară, fiind structurată logic pe patru capitole teoretice și aplicative, urmate de concluzii și direcții viitoare. Lucrarea își propune să identifice cea mai performantă metodă de predicție aplicabilă în România, contribuind totodată la dezvoltarea cercetării în domeniul evaluării automatizate a proprietăților imobiliare.

Capitolul 1 – Evaluarea activelor în diferite scopuri

Evaluarea proprietăților imobiliare s-a dezvoltat accelerat în România ca urmare a creșterii pieței imobiliare și a cererii pentru servicii specializate. Această activitate profesională, integrată în sfera serviciilor de tip consultativ, presupune competență avansată, respectarea unui cod etic și aplicarea unor standarde internaționale.

1.1 Delimitări generale

1.1.1 Istoricul profesiei de evaluator

Originile profesiei se regăsesc în Marea Britanie (1834), iar dezvoltarea sa internațională a fost marcată de formarea unor organizații cheie precum RICS (UK), Appraisal Institute (SUA) și IVSC (la nivel global). În Europa, standardizarea a fost consolidată prin TEGoVA. În România, profesia

s-a oficializat după 1990, prin înființarea ANEVAR, care a devenit o instituție de utilitate publică ce reglementează și sprijină activitatea evaluatorilor autorizați.

1.1.2 Noțiuni-cheie: preț, cost și valoare

Evaluarea implică distincția între:

- **Preț** – suma efectivă tranzacționată;
- **Cost** – cheltuiala necesară pentru realizarea unui activ;
- **Valoare** – percepția de piață privind utilitatea unui activ.

Acestea sunt influențate de factori precum utilitatea, raritatea, preferința și capacitatea de cumpărare, care reflectă echilibrul dintre cerere și ofertă.

1.1.3 Obiectul evaluării

Evaluarea este o activitate profesională fundamentată pe judecata expertului și exprimată într-un raport oficial. Se aplică asupra unor categorii distincte de active:

- Proprietăți imobiliare;
- Bunuri mobile;
- Întreprinderi și participații;
- Instrumente financiare.

Lucrarea se concentrează asupra **proprietăților imobiliare rezidențiale**, inclusiv asupra drepturilor parțiale asociate (ex: folosință, închiriere, ipotecă).

1.2 Abordări, metode și tehnici de evaluare

Evaluarea proprietăților imobiliare presupune aplicarea uneia sau mai multora dintre cele trei abordări consacrate internațional: abordarea prin **cost**, prin **pieță** și prin **venit**. Alegerea metodei optime depinde de tipul proprietății și de contextul pieței.

1.2.1 Abordarea prin piață

Această metodă se bazează pe analiza comparativă a unor tranzacții recente cu proprietăți similare. Este considerată metoda de referință pentru evaluările rezidențiale, întrucât reflectă comportamentul efectiv al pieței. Procesul implică identificarea unor active comparabile, ajustarea diferențelor dintre acestea și proprietatea evaluată, și utilizarea unor unități de măsură adecvate (ex. €/mp sau €/hectar).

Elemente-cheie în analiză includ localizarea, starea clădirii, drepturile de proprietate, regimul juridic și condițiile de piață. Comparabilitatea este validată prin selecția tranzacțiilor recente, preferabil din aceeași zonă, și prin ajustări explicite ale prețurilor, exprimate în valoare absolută sau procentual.

1.2.1.1 Factorii care influențează prețurile

Prețul unei proprietăți este afectat de o gamă largă de variabile – de la caracteristicile fizice ale terenului și construcției, la reglementări urbanistice, utilități, poziționare și accesibilitate. Un accent special se pune pe diferențele între tranzacțiile analizate și proprietatea evaluată, iar aceste variații sunt corectate sistematic.

1.2.1.2 Tehnici de corecție

Corecțiile se pot realiza prin metode cantitative (analize pe perechi de date, regresii, tendințe statistice) sau prin metode descriptive (comparare relativă, clasificări, interviuri cu experți). Fiecare tehnică ajută la ajustarea valorilor în funcție de trăsături specifice ale fiecărei proprietăți.

1.2.1.3 Conformitatea cu standardele din România

Metoda comparației directe este în acord cu Standardele de Evaluare a Bunurilor (SEV) și este larg utilizată pentru evaluarea drepturilor imobiliare. Se impune folosirea unor unități de comparație relevante și acceptate în piață, precum €/mp sau €/cameră, și aplicarea coerentă a acestora.

Evaluatorul trebuie să fie atent la diferențele dintre proprietăți, să acorde prioritate tranzacțiilor recente și să descrie clar în raport toate ajustările făcute. În lipsa unor comparații directe, se pot folosi și metode indirecte (raport, deducție), însă cu rezerve privind precizia.

1.2 Abordări, metode și tehnici de evaluare

În practica evaluării imobiliare, sunt utilizate trei mari abordări metodologice: **prin piață, prin venit și prin cost**. Fiecare se bazează pe principii economice solide și este aleasă în funcție de tipul proprietății și disponibilitatea datelor.

1.2.1 Abordarea prin piață

Această metodă compară activul evaluat cu tranzacții recente ale unor proprietăți similare. Evaluatorul selectează unități de măsură relevante (€/mp, €/cameră etc.), identifică diferențele și aplică ajustări specifice pentru a reflecta caracteristicile proprietății analizate.

Comparabilitatea este esențială, iar metodele utilizate includ analize cantitative (regresii, perechi de date, grafică) și descriptive (clasificări comparative, interviuri).

Este abordarea dominantă în evaluarea proprietăților rezidențiale, fiind conformă cu standardele românești și internaționale. Precizia acesteia depinde de calitatea și recența datelor de piață.

1.2.2 Abordarea prin venit

Se aplică în cazul proprietăților generatoare de profit (ex. clădiri de birouri, hoteluri), prin conversia fluxurilor viitoare de venit în valoare actuală. Există două metode principale:

- **Capitalizarea directă** – simplă, bazată pe un venit anual stabil și rata de capitalizare.
- **Discounted Cash Flow (DCF)** – actualizează venituri multiple pe un orizont de timp, fiind mai complexă și adaptată previziunilor dinamice.

Această abordare presupune o înțelegere profundă a pieței de închiriere, a riscurilor și a anticipărilor investitorilor, fiind susținută de principiile cererii și anticipării.

1.2.3 Abordarea prin cost

Această metodă estimează valoarea unui activ pornind de la costul de reconstrucție sau înlocuire, din care se scad deprecierile fizice, funcționale și externe. Este preferată în cazul construcțiilor noi, specializate sau în lipsa unui volum suficient de tranzacții.

Evaluarea presupune parcurgerea următorilor pași:

- calculul costului de nou (prin deviz, cost unitar sau comparație);
- determinarea deprecierilor (recuperabile sau nerecuperabile);
- estimarea valorii nete.

Se disting costuri directe (materiale, manoperă, profit antreprenor) și indirecte (onorarii, studii, taxe, marketing). Acestea trebuie complet și corect integrate în analiza finală.

1.3 Scopul și utilitatea evaluării

Evaluarea imobiliară are drept scop estimarea unei valori asociate unui drept de proprietate, în funcție de nevoile concrete ale beneficiarului. Aceasta poate servi unei game largi de obiective – de la tranzacții comerciale și garantare bancară, la stabilirea impozitelor sau întocmirea situațiilor contabile.

Scopul evaluării este determinat de solicitant și vizează luarea unei decizii informate privind activul imobiliar. Deși forma raportului poate varia în funcție de utilizare, valoarea estimată rămâne neschimbată, fiind rezultatul unui proces standardizat și obiectiv.

În această lucrare sunt detaliate trei dintre cele mai frecvente scopuri ale evaluării:

1. Evaluarea pentru vânzare/cumpărare

Este cea mai uzuală formă de evaluare, având rolul de a sprijini negocierea unui preț corect între părți. Modelele automatizate (AVM) pot eficientiza acest proces, oferind estimări rapide și consistente, bazate pe date istorice și predictive. Ele contribuie la eliminarea subiectivității și la luarea unor decizii transparente atât pentru cumpărători, cât și pentru vânzători.

2. Evaluarea pentru garantare bancară

Utilizată de bănci pentru stabilirea valorii colateralului, această formă de evaluare este crucială în procesul de acordare a creditelor. Integrarea învățării automate permite obținerea unor estimări rapide, precise și standardizate, reducând riscurile de supra- sau subevaluare. Pentru bănci, această eficientizare sporește gestionarea riscurilor și optimizează procesul decizional.

3. Evaluarea pentru impozitare

Are rolul de a determina valoarea fiscală a imobilelor, influențând direct nivelul taxelor datorate. Evaluarea este reglementată legal și trebuie realizată periodic, în special în cazul clădirilor nerezidențiale. Modelele predictive pot susține un sistem fiscal mai echitabil, prin estimări uniforme, transparente și automatizate, contribuind la o colectare eficientă a veniturilor publice.

În toate cele trei cazuri, automatizarea procesului de evaluare – prin utilizarea algoritmilor de învățare automată – îmbunătățește acuratețea, reduce costurile și oferă suport decizional robust pentru toți actorii implicați: proprietari, instituții financiare, autorități și dezvoltatori.

1.4 Concluziile Capitolului 1

Evaluarea proprietăților imobiliare reprezintă un pilon esențial al economiei moderne, consolidat prin standarde profesionale, raționament expert și o infrastructură instituțională solidă. În România, această profesie s-a dezvoltat accelerat după 1990, fiind reglementată de ANEVAR și aliniată cu organisme internaționale precum IVSC și TEGoVA.

Procesul de evaluare nu se rezumă la calcule matematice, ci implică judecată profesională și cunoașterea aprofundată a pieței. Cele trei abordări fundamentale – prin **piață**, **venit** și **cost** – sunt aplicate în funcție de scopul evaluării și caracteristicile activului, abordarea prin piață fiind dominantă în segmentul rezidențial.

Valoarea estimată devine baza unor decizii economice importante: tranzacții de vânzare, garantare bancară, fiscalitate, raportare contabilă. Prin urmare, evaluarea răspunde nevoilor nu doar ale persoanelor fizice sau companiilor, ci și ale instituțiilor financiare, autorităților fiscale și investitorilor.

În special, **evaluările pentru vânzare-cumpărare, garantare bancară și impozitare** sunt de maximă relevanță în contextul actual. Acestea pot fi optimizate semnificativ prin integrarea învățării automate, care permite analizarea masivă a datelor și oferă estimări rapide, obiective și scalabile. Modelele predictive reduc incertitudinea, elimină subiectivismul și sprijină transparența în luarea deciziilor.

Acest capitol oferă cadrul teoretic necesar pentru a înțelege nu doar esența procesului de evaluare, ci și justificarea utilizării tehnologiilor avansate în acest domeniu. Astfel, el fundamentează direcția cercetării ulterioare – automatizarea procesului de estimare a valorii imobiliare – prin clarificarea conceptelor, scopurilor, metodelor și standardelor de evaluare.

Capitolul 2 – Analiza literaturii privind evaluările automate

Acest capitol explorează evoluția, aplicabilitatea și validarea modelelor automatizate de evaluare a proprietăților (AVM), accentuând sinergia dintre metodele clasice de evaluare și algoritmi de învățare automată. AVM-urile folosesc seturi extinse de date și modele predictive pentru a estima valoarea de piață a imobilelor, fără intervenție umană directă. Această abordare optimizează acuratețea, eficiența și obiectivitatea procesului de evaluare, devenind tot mai utilizată în contextul digitalizării pieței imobiliare.

2.1 Metodologia analizei literaturii

Pentru analiza sistematică, au fost selectate 203 articole din peste 20.000 identificate în baze de date științifice precum Scopus, Web of Science și Science Direct. Selecția a fost realizată pe baza unor criterii riguroase: relevanță tematică, impact științific, accesibilitate completă și validare metodologică. Accentul s-a pus pe lucrări care analizează direct AVM-urile, folosesc metode cantitative/statistice și explorează implicațiile acestora în evaluarea imobiliară.

Teme cheie identificate în literatură

Cercetarea a identificat mai multe direcții principale:

- Definirea și clasificarea AVM-urilor;
- Tipologia datelor și variabilelor utilizate;

- Modele statistice aplicate (regresii, rețele neuronale, metode hibride);
- Utilitatea AVM-urilor în contexte geografice și instituționale variate;
- Analiza SWOT și impactul reglementărilor asupra adoptării AVM-urilor.

Tabelul central al capitolului rezumă 14 articole reprezentative, detaliind scopul cercetării, metodologia, variabilele folosite, dimensiunea eșantionului și concluziile relevante pentru prezenta lucrare.

Integrarea modelelor cantitative și calitative

Studiile cantitative analizează performanța și precizia AVM-urilor pe baze de date reale, în timp ce studiile calitative oferă un cadru conceptual și metodologic pentru înțelegerea contextului aplicării lor. Împreună, aceste abordări sprijină dezvoltarea unui model propriu adaptat pieței românești.

Contribuții și relevanță

Capitolul fundamentează necesitatea aplicării modelelor AVM în evaluarea imobiliară, oferind un cadru de referință solid pentru definirea metodologiei proprii în cercetare. Prin analiza bibliometrică și evaluarea critică a literaturii, sunt identificate cele mai relevante surse, metodologii și variabile independente pentru construirea unui model predictiv robust.

Subcapitolul 2.2: Analiza bibliometrică a literaturii

Subcapitolul 2.2 propune o investigație bibliometrică detaliată asupra literaturii de specialitate referitoare la modelele automate de evaluare (AVM), cu scopul de a evidenția evoluția cercetării, centrele academice implicate și temele dominante. Analiza se bazează pe baza de date Web of Science și include 318 articole publicate între 1998 și 2024. Deși volumul total este relativ modest, intensificarea cercetării este evidentă, cu un vârf de publicări în anul 2017 și un maxim de citări în 2024.

Studiile sunt predominant distribuite în domeniile inteligenței artificiale și informaticii, dar și în economie, finanțe, studii urbane sau inginerie. Tehnologiile emergente, precum învățarea automată, regresia, arborii de decizie, rețelele neuronale sau metodele ensemble (ex. random

forest, XGBoost), sunt frecvent asociate cu AVM-urile, reflectând direcțiile principale de dezvoltare.

Distribuția geografică arată o concentrare a cercetării în țări cu economii avansate (SUA, Marea Britanie, China, Germania), dar și o participare activă din partea unor state europene și asiatice emergente. Jurnalele de specialitate cu impact ridicat în acest domeniu provin din arii variate – de la evaluare imobiliară și finanțe, la inteligență artificială – sugerând o abordare interdisciplinară consolidată.

Analiza terminologică realizată cu VOSviewer scoate în evidență legătura strânsă între AVM-uri și concepte precum prețul imobilelor, evaluarea în masă, piața imobiliară și garantarea creditelor. Aceste conexiuni validează relevanța subiectului tezei și susțin integrarea modelelor de învățare automată în evaluarea imobiliară contemporană.

În concluzie, analiza bibliometrică oferă o bază solidă pentru aprofundarea ulterioară a cercetării, relevând atât progresele metodologice, cât și nevoia de consolidare a aplicabilității practice și a transparenței modelelor AVM. Subcapitolul pregătește astfel terenul pentru o investigație calitativă sistematică, ce va fi realizată în secțiunea următoare.

Subcapitolul 2.3 explorează literatura de specialitate printr-o analiză sistematică axată pe studiile empirice și metodele utilizate în construcția modelelor automate de evaluare (AVM). Spre deosebire de abordarea bibliometrică, această analiză urmărește aprofundarea calitativă a contribuțiilor academice, concentrându-se pe cele mai relevante 50 de articole identificate în Web of Science și Scopus, în funcție de citări și relevanță tematică.

Sinteza evidențiază aplicabilitatea AVM-urilor în diverse contexte: evaluări rezidențiale, estimări fiscale, împrumuturi garantate și analize de piață. De asemenea, sunt analizate metodele statistice și algoritmi de machine learning utilizați frecvent – regresia liniară, regresia hedonică, arborii de decizie, pădurile aleatorii, rețelele neuronale și XGBoost – în funcție de scopul cercetărilor și calitatea datelor disponibile.

Majoritatea studiilor analizate au evidențiat avantaje semnificative ale modelelor AVM în ceea ce privește eficiența și acuratețea, însă au remarcat și limitări precum lipsa transparenței, dificultatea

în interpretarea rezultatelor și dependența de calitatea datelor. În plus, unele cercetări pun accent pe integrarea expertizei umane și a tehnicilor hibride pentru a spori robustețea estimărilor.

Analiza confirmă caracterul multidisciplinar al cercetării în domeniul AVM-urilor, cu contribuții din domenii precum inteligența artificială, economie, urbanism și inginerie. Totodată, se conturează nevoia de standardizare a metodologiilor și a seturilor de date utilizate, pentru a facilita comparabilitatea și reproducerea rezultatelor.

În concluzie, analiza sistematică pune în evidență progresele semnificative realizate în proiectarea și testarea AVM-urilor, dar și necesitatea continuării cercetărilor în direcția integrării datelor spațiale, explicabilității modelelor și adaptabilității la condiții de piață diferite.

Concluziile Capitolului 2

Capitolul 2 oferă o imagine comprehensivă asupra modului în care modelele automate de evaluare (AVM) au fost abordate în literatura științifică, subliniind tranziția de la metodele tradiționale la cele automatizate. Analiza literaturii, bazată pe surse de prestigiu, a evidențiat utilizarea frecventă a AVM în piețele imobiliare avansate și consolidarea acestora ca instrumente utile în contextul digitalizării evaluării.

Studiile revizuite confirmă eficiența AVM în termeni de viteză, costuri reduse și capacitate de analiză în masă, dar atrag atenția și asupra limitărilor acestora: dependența de date de calitate, dificultatea de adaptare în piețele mai puțin lichide și provocările legate de transparența algoritmilor. Se remarcă totodată nevoia de integrare hibridă între modelele automate și expertiza umană, pentru a spori robustețea și validitatea estimărilor.

Analiza evidențiază că AVM-urile sunt cel mai eficient aplicabile în evaluările de masă, în special pentru impozitare, garantare bancară și analiza riscurilor. Studiile recente pun accent pe optimizarea variabilelor de intrare, evitarea suprapotrivirii și calibrarea constantă a modelelor. În același timp, abordările hibride și utilizarea inteligenței artificiale susțin progresul metodologic al AVM-urilor.

Tabelul de sinteză inclus în capitol prezintă comparativ principalele beneficii și limitări ale AVM, consolidând concluzia că aceste modele reprezintă un sprijin valoros în procesul decizional, dar nu pot substitui complet judecata profesională. Astfel, AVM-urile se conturează ca o soluție complementară, cu potențial semnificativ de extindere în contextul creșterii volumului de date și a nevoii de eficiență în evaluarea proprietăților.

Capitolul 3: Modele de învățare automată și aplicabilitatea acestora în evaluarea proprietăților imobiliare

Capitolul 3 explorează aplicabilitatea modelelor de învățare automată (ML) în evaluarea proprietăților imobiliare, evidențiind capacitatea acestora de a analiza volume mari de date și de a identifica relații complexe, neliniare, care scapă metodelor tradiționale. În contextul unei piețe imobiliare tot mai dinamice, ML se conturează drept o soluție eficientă pentru creșterea acurateței și obiectivității în estimarea valorilor imobilelor.

Sunt prezentate principalele modele ML utilizate în evaluare: regresia liniară, arborii de decizie, pădurile aleatorii, rețelele neuronale artificiale și Extreme Gradient Boosting. Fiecare tehnică este analizată în funcție de principiul de funcționare, avantajele și limitările sale, demonstrând aplicabilitatea lor în diverse contexte de piață. Performanța acestor modele este evaluată prin indicatori specifici – RMSE, MAE, R^2 , F-statistic și P-value – pentru a asigura validitatea predicțiilor.

RStudio este menționat ca mediu preferat pentru implementarea și testarea acestor modele, oferind flexibilitate în prelucrarea datelor, selecția hiperparametrilor și interpretarea rezultatelor.

Subcapitolul 3.1 analizează extensia învățării automate către sisteme integrate de inteligență artificială (IA), clasificând domeniile acesteia (învățare, interfață naturală, percepție, robotică etc.) și identificând aplicațiile relevante pentru evaluarea imobiliară – în special sistemele expert, rețelele neuronale și agenții inteligenți. Se subliniază faptul că IA nu doar sprijină, ci și extinde potențialul predictiv al ML, consolidând obiectivitatea și eficiența procesului evaluativ.

Tabelele incluse oferă o imagine de ansamblu asupra clasificării domeniilor IA (Tabelul 7) și asupra implementării practice a modelelor ML în diverse industrii (Tabelul 8), de la energie și

agricultură, până la piețe financiare și sănătate. Această transversalitate confirmă versatilitatea acestor modele și potențialul de transfer în domeniul evaluării imobiliare.

În concluzie, învățarea automată reprezintă un instrument inovator și validat științific pentru estimarea valorii proprietăților, susținând transformarea digitală a procesului de evaluare. Modelele ML permit o analiză aprofundată, reduc erorile umane și susțin luarea deciziilor în mod transparent și predictiv, deschizând noi perspective pentru modernizarea pieței imobiliare.

3.2: Modele de învățare automată utilizate în evaluarea proprietăților imobiliare

Subcapitolul 3.2 analizează comparativ cele mai utilizate modele de învățare automată aplicabile în evaluarea proprietăților imobiliare, punând accent pe acuratețea estimărilor, robustețea și eficiența computațională. Modelele examinate includ: regresia liniară, arborii de decizie, pădurile aleatorii, rețelele neuronale artificiale și Extreme Gradient Boosting (XGBoost).

Fiecare model este prezentat din perspectiva principiilor de funcționare, a avantajelor oferite și a limitărilor specifice. Regresia liniară se remarcă prin simplitate și interpretabilitate, fiind potrivită pentru relații directe între variabile. Arborii de decizie permit modelarea relațiilor complexe și oferă claritate decizională, însă pot deveni instabili. Pădurile aleatorii cresc stabilitatea și precizia prin combinarea mai multor arbori. Rețelele neuronale se adaptează excelent la pattern-uri nelineare, dar necesită resurse semnificative și un set de date consistent. XGBoost, model de boosting avansat, se dovedește foarte eficient în reducerea erorilor, fiind capabil să trateze outlierii și variabilele complexe cu performanță ridicată.

Evaluarea modelelor se bazează pe indicatori de performanță precum eroarea medie pătratică (RMSE), eroarea medie absolută (MAE) și coeficientul de determinare (R^2), oferind un cadru obiectiv de comparație. Studiile de caz și experimentele realizate arată că XGBoost oferă cel mai bun echilibru între acuratețea predicțiilor și eficiența în procesarea datelor, în timp ce pădurile aleatorii oferă robustețe și generalizare bună în condiții diverse de piață.

În concluzie, selecția modelului optim trebuie să țină cont de specificul pieței imobiliare, volumul și calitatea datelor disponibile, precum și scopul evaluării. Modelele avansate, precum XGBoost și rețelele neuronale, sunt recomandate pentru predicții de înaltă precizie, în timp ce modelele mai

simple pot fi preferate în contextul unei interpretabilități ridicate sau a unui volum restrâns de date. Astfel, învățarea automată se dovedește un instrument versatil și scalabil în procesul modern de evaluare imobiliară.

3.4: Evaluarea performanței modelelor de învățare automată

Subcapitolul 3.4 evidențiază importanța evaluării riguroase a modelelor de învățare automată utilizate în estimarea valorii proprietăților imobiliare. Analiza performanței se bazează pe o serie de metrice statistice esențiale, care permit compararea obiectivă a preciziei algoritmilor și fundamentarea alegerii celui mai eficient model.

Cele mai relevante instrumente de validare sunt: RMSE (Root Mean Squared Error), care penalizează erorile mari și oferă o imagine de ansamblu asupra abaterilor de la valorile reale; MAE (Mean Absolute Error), care exprimă eroarea medie în termeni absoluți și este mai puțin sensibil la outlieri; R^2 (coeficientul de determinare), care reflectă proporția din variabilitatea prețurilor explicată de model; și indicatorii F-statistic și P-value, esențiali pentru testarea semnificației statistice a relațiilor dintre variabile.

RMSE este util în scenarii unde erorile mari sunt costisitoare, în timp ce MAE oferă o interpretare intuitivă, mai apropiată de realitatea operațională. Compararea valorilor acestor două metrice poate semnaliza prezența valorilor extreme. R^2 ajută la înțelegerea capacității explicative a modelului, dar poate supraestima performanța în cazurile de supraajustare. F-statistic și P-value evaluează validitatea globală a modelului și contribuția fiecărei variabile la precizia predicției.

Aceste evaluări sunt aplicate în medii software specializate, precum RStudio, care oferă un cadru robust pentru analiza performanței predictive. În contextul studiului, utilizarea acestor metrice a permis compararea obiectivă a modelelor testate, susținând alegerea algoritmilor cu cea mai mare precizie în estimarea valorii proprietăților.

3.5 Modul de lucru în programul informatic RStudio – Rezumat original

RStudio este o platformă integrată, bazată pe limbajul R, destinată prelucrării și analizei avansate a datelor, utilizată extensiv în modelarea predictivă. Interfața sa intuitivă permite editarea scripturilor, rularea codului, vizualizarea grafică a datelor și gestionarea pachetelor necesare.

Importul datelor se realizează facil, prin funcții precum `read.csv()` sau `read_excel()`, urmat de etape esențiale de curățare și transformare, frecvent realizate cu ajutorul pachetului `dplyr`. Modelarea statistică poate fi efectuată prin funcții precum `lm()` pentru regresie liniară, completată de teste de semnificație statistică relevante.

Un avantaj major al platformei este capacitatea sa de vizualizare prin `ggplot2`, care permite generarea de grafice personalizabile – utile pentru explorarea relațiilor dintre variabile. De asemenea, RStudio susține redactarea de rapoarte interactive și reproductibile în formate PDF/HTML, prin `rmarkdown`, combinând codul, textul și rezultatele analitice într-un document unic.

În contextul lucrării de față, RStudio este utilizat pentru implementarea și validarea modelelor de învățare automată destinate estimării valorii imobiliare, fiind esențial pentru aplicarea metodelor statistice și evaluarea performanței algoritmilor prin indicatori-cheie precum RMSE, MAE, R^2 , F-statistic și P-value.

Concluzii Capitol 3

Capitolul analizează aplicabilitatea modelelor de învățare automată în estimarea valorii proprietăților imobiliare, subliniind avantajele acestora față de metodele tradiționale, mai ales în contexte complexe și dinamice. Învățarea automată permite identificarea relațiilor nelineare dintre variabile și realizarea de predicții precise, obiective și reproductibile.

În prima parte, este discutată integrarea acestor tehnologii în evaluarea proprietăților, inclusiv cadrul mai larg al inteligenței artificiale și subdomeniile relevante (ex: sisteme expert, rețele neuronale, agenți inteligenți). Modelele utilizate frecvent sunt regresia liniară, arborii de decizie, pădurile aleatorii, rețelele neuronale artificiale și Extreme Gradient Boosting – fiecare prezentând

avantaje distincte, dar și limitări privind interpretabilitatea, consumul computațional sau riscul de supraînvățare.

Capitolul include o analiză a performanței acestor modele folosind indicatori statistici precum RMSE, MAE, R^2 , F-statistic și P-value, esențiali pentru validarea rezultatelor. Este evidențiată importanța echilibrului între precizia estimărilor și robustețea.

Totodată, este descris modul de implementare practică în RStudio, platforma utilizată pentru dezvoltarea și testarea modelelor, cu accent pe importul datelor, curățarea, vizualizarea și generarea rapoartelor.

Concluziile subliniază că niciun model nu este universal superior, alegerea optimă depinzând de contextul evaluării și de natura datelor disponibile. Capitolul pregătește trecerea către studiul de caz, care va demonstra aplicabilitatea practică a acestor concepte în analiza prețurilor imobiliare reale.

Capitolul 4 – Evaluarea performanțelor modelelor de învățare automată în predicția prețurilor proprietăților imobiliare de tip rezidențial

Într-un context dominat de creșterea volumului de date și de progresele rapide în domeniul învățării automate, acest capitol analizează aplicarea practică a cinci tehnici predictive în estimarea valorii apartamentelor rezidențiale din Cluj-Napoca. Scopul principal este identificarea metodei cu cea mai mare acuratețe, adaptată particularităților pieței locale.

Studiul comparativ include: regresia liniară, arborii de regresie, pădurile aleatorii, rețelele neuronale artificiale (ANN) și XGBoost. Fiecare tehnică este evaluată în funcție de performanța predictivă, complexitatea implementării și robustețea la variațiile din date. Procesul de construire a modelelor implică selecția variabilelor, optimizarea hiperparametrilor și validarea riguroasă prin cross-validation. Metricile utilizate – RMSE și MAE – permit o evaluare echilibrată între precizia estimărilor și rezistența la valori aberante.

Prin această abordare, capitolul fundamentează necesitatea trecerii de la metode tradiționale la soluții automatizate, adaptate la dinamica unei piețe imobiliare complexe. Implementarea acestor

modele predictive vizează obținerea unor estimări rapide, obiective și reproductibile, contribuind la profesionalizarea procesului de evaluare imobiliară și la optimizarea deciziilor economice în sectorul rezidențial.

4.2 – Setul de date

Pentru a fundamenta modelele de învățare automată destinate predicției prețurilor locuințelor, cercetarea a utilizat un set de date riguros construit, reprezentativ pentru piața imobiliară din Cluj-Napoca. Acest oraș a fost ales datorită poziției sale de lider în sectorul rezidențial din România, evidențiindu-se prin prețuri ridicate, cerere intensă și venituri medii peste media națională.

Setul de date este compus din 773 de apartamente cu una până la patru camere, selectate din două surse complementare: platforma Argus, destinată profesioniștilor din domeniul imobiliar, și rapoarte de evaluare realizate de firma Napoca Business S.R.L. Procesul de selecție a inclus doar anunțurile care furnizau informații complete și verificabile, fiind eliminate ofertele incomplete, redundante sau nesigure.

Perioada analizată acoperă anii 2019–2020, iar datele au fost completate și validat manual de către autor, evaluator autorizat, inclusiv prin apeluri telefonice către vânzători. Prin acest demers s-a asigurat un grad ridicat de acuratețe și fidelitate față de realitatea pieței. Efortul de prelucrare a dus la excluderea a peste 90% din anunțurile inițiale, pentru a garanta integritatea și relevanța bazei finale.

Acest set de date oferă o bază solidă pentru calibrări statistice și testarea performanței modelelor predictive, permițând o explorare nuanțată a relațiilor dintre variabile și prețurile imobiliare. Subcapitolul următor va detalia aceste variabile și modul în care contribuie la construcția modelelor de predicție.

4.3 Variabilele

Studiul propus se bazează pe un set complex de **36 de variabile**, împărțite în două categorii principale: **cantitative** (ex. suprafața utilă, numărul de camere, distanța față de centru) și **calitative (categoriale)** (ex. tipul imobilului, facilități, vecinătăți). Aceste variabile reflectă practicile curente

în evaluarea imobiliară tradițională și au fost codificate corespunzător în RStudio pentru o integrare optimă în modelele de învățare automată.

Pentru zona analizată – **Cluj-Napoca** – selecția variabilelor a vizat acoperirea caracteristicilor esențiale ale apartamentelor: localizare, structură, compartimentare, facilități și accesibilitate urbană. Informațiile au fost colectate din anunțurile platformei Argus, prelucrate și agregate pe cartiere, iar variabilele au fost validate inclusiv cu suportul unor instrumente precum Google Maps.

Printre variabilele cheie se numără: **cartierul, suprafața, tipul construcției, prețul, distanța față de centru**, precum și proximitatea față de puncte de interes urban (instituții, spații verzi, centre comerciale). Au fost integrate și variabile de confort (balcon, lift, parcare, boxă), precum și cele privind **eficiența energetică și nivelul de poluare**.

O atenție specială a fost acordată variabilei „**concentrație**”, care cuantifică densitatea construcțiilor pe o rază de 500 m², fiind transformată în factor ordinal (scăzută, medie, ridicată). De asemenea, s-au analizat preferințele cumpărătorilor privind **numărul de camere, tipul compartimentării, tipul construcției (nou/vechi) și prezența locurilor de parcare**, prin reprezentări grafice generate în RStudio.

Datele colectate și structura variabilelor contribuie decisiv la fundamentarea unor modele predictive robuste, reflectând dinamica reală a pieței imobiliare din Cluj-Napoca. Acest cadru permite identificarea celor mai influenți factori în determinarea valorii apartamentelor și susține procesul decizional în evaluare și investiții imobiliare.

4.4 Rezultatele obținute și interpretarea acestora

Analiza comparativă a celor cinci modele de învățare automată aplicate pentru predicția prețurilor apartamentelor din Cluj-Napoca a evidențiat diferențe semnificative în performanță. Modelele testate – regresia liniară, arborii de decizie, pădurile aleatorii, rețelele neuronale artificiale și XGBoost – au fost evaluate pe baza metricalor RMSE și MAE, utilizând validare încrucișată (cross-validation) pentru robustețe.

Modelul XGBoost s-a remarcat drept cel mai performant, obținând cele mai mici valori ale erorilor de predicție. Acesta a demonstrat o capacitate superioară de a surprinde relațiile complexe și nelineare dintre variabilele analizate. Pe locurile următoare s-au clasat rețelele neuronale artificiale (ANN) și pădurile aleatorii, ambele cu rezultate bune, dar cu un consum mai mare de resurse computaționale.

Regresia liniară, deși interpretabilă și ușor de implementat, a avut cele mai slabe rezultate, ceea ce confirmă limitările acestei metode în contextul relațiilor nelineare și al complexității pieței. **Arborii de decizie**, simpli și explicabili, au avut performanță medie, dar au fost afectați de tendința de supraînvățare.

Rezultatele obținute susțin ideea că modelele avansate de tip boosting sau rețele neuronale oferă o estimare mai precisă a prețurilor imobiliare, în special pe piețele dinamice precum cea din Cluj-Napoca. Totodată, analiza relevă că alegerea modelului optim trebuie făcută în funcție de scopul evaluării, complexitatea variabilelor și resursele disponibile.

4.4.2 Arborii de decizie

Modelul bazat pe **arbori de decizie (CART)** a fost implementat pentru estimarea prețului apartamentelor, fiind calibrat prin tehnica **grid search** pentru optimizarea hiperparametrilor. Au fost testate peste 21.000 de combinații, fiecare evaluată prin validare încrucișată (5-fold cross-validation), utilizând metricile RMSE, MAE și R^2 .

Cei mai importanți hiperparametri ajustați au inclus adâncimea maximă a arborelui, mărimea minimă a nodurilor terminale (MinBucket), numărul minim de observații pentru divizare (MinSplit) și parametrul de complexitate. Rezultatul optim a fost obținut pentru un arbore cu adâncime 6, MinSplit = 10, MinBucket = 19 și CP = 0.001, generând un RMSE de **18.052,18** și un R^2 de **0.732**.

Graficul arborelui a arătat că **Suprafața apartamentului** este cel mai influent factor în divizare, urmat de variabile precum **Cartier**, **Număr camere**, **Izolare termică** și **Loc de parcare**. De asemenea, s-a realizat o ierarhizare a variabilelor în funcție de contribuția la reducerea varianței, confirmând rolul esențial al dimensiunii și compartimentării în determinarea prețului.

Modelul s-a dovedit eficient în captarea relațiilor non-liniare, însă necesită atenție sporită la reglarea hiperparametrilor pentru a evita supraînvățarea. Comparativ cu regresia liniară, arborii de decizie oferă o interpretare vizuală clară și o flexibilitate sporită în analiza datelor imobiliare. Totuși, limitările privind robustețea au condus la necesitatea explorării unei metode superioare – **pădurile aleatorii**, analizate în subcapitolul următor.

4.4.3 Păduri aleatorii

Metoda **pădurilor aleatorii**, propusă ca a treia tehnică în cadrul studiului de caz, constă în agregarea predicțiilor generate de un ansamblu de arbori de decizie, pentru a obține o estimare robustă a prețului apartamentelor. Implementarea a fost realizată în RStudio, folosind librăriile `caret`, `randomForest` și `dplyr`, împreună cu o validare încrucișată pe cinci subseturi și o optimizare a hiperparametrilor prin grid search.

Parametrii calibrați au inclus: **numărul de arbori (Ntree)**, **numărul de variabile selectate la fiecare divizare (Mtry)** și **dimensiunea minimă a nodurilor terminale (Nodesize)**. După testarea a 345 de combinații, configurația optimă a fost identificată la:

- $Ntree = 179$
- $Mtry = 22$
- $Nodesize = 9$
- $RMSE = 17.532,62$
- $R^2 = 81,96\%$

Acești indicatori reflectă o **precizie ridicată și o capacitate solidă de generalizare**, superioare modelelor testate anterior. Importanța variabilelor a fost evaluată prin metoda *Mean Decrease in Accuracy*, care a evidențiat că **Suprafața**, **Numărul de camere**, **Numărul de etaje**, **Distanța față de centru** și **Cartierul** sunt determinanți principali în predicția prețurilor.

Graficul de învățare a confirmat că după aproximativ 200 de arbori, performanța modelului se stabilizează, indicând atingerea punctului optim fără riscul de supraînvățare. De asemenea, analiza importanței variabilelor a arătat că infrastructura urbană (transport public, instituții sanitare) are

un impact moderat, în timp ce factori precum apropierea de centre comerciale sau instituții religioase au avut o influență neglijabilă.

Prin compararea cu modelele anterioare, pădurile aleatorii s-au dovedit **mai performante decât regresia liniară și arborii de decizie simpli**, oferind o estimare stabilă, interpretabilă și eficientă pentru valoarea apartamentelor din Cluj-Napoca.

4.4.4 Rețele Neuronale Artificiale

În cadrul studiului de caz, **rețelele neuronale artificiale dense (ANN)** au fost utilizate pentru a modela relațiile complexe și non-liniare dintre caracteristicile apartamentelor și prețul acestora. Implementarea s-a realizat în Google Colab, cu ajutorul bibliotecii Keras și limbajului R, iar pregătirea datelor a inclus normalizarea variabilelor numerice și transformarea celor categorice prin one-hot encoding.

Modelul a fost antrenat pe baza unui proces riguros de optimizare, implicând:

- alegerea numărului optim de **epoci** (identificat ca fiind 173),
- utilizarea algoritmului **RMSprop** cu reglaj fin al hiperparametrilor **learning rate** și **momentum**,
- testarea diferitelor **arhitecturi** de rețea, cu unul sau două straturi ascunse, fiecare având între 32 și 512 neuroni.

Performanța modelelor a fost evaluată cu ajutorul validării încrucișate (5-fold), pe setul de antrenament, folosind metricile RMSE și MAE. Arhitectura optimă a fost formată din **două straturi ascunse (64 și 32 de neuroni)**, iar combinația de hiperparametri selectată (learning rate = 0.005, momentum = 0.95) a generat un RMSE de **16.940,07**, cel mai scăzut din toate modelele analizate până acum.

Graficul de învățare a demonstrat o reducere progresivă a erorii pe seturile de antrenament și validare, validând eficiența rețelei în procesul de învățare. Divergențele observate la un număr prea mare de epoci au semnalat riscul de supraînvățare, fiind evitat prin calibrarea atentă a parametrilor.

Modelul ANN a depășit performanțele regresiei liniare, arborilor de decizie și pădurilor aleatorii, oferind o **precizie superioară în estimarea prețurilor imobiliare**. Acest rezultat confirmă potențialul ridicat al rețelelor neuronale dense în captarea tiparelor complexe din piața imobiliară și în susținerea deciziilor automatizate în evaluare.

4.4.5 Extreme Gradient Boosting (XGBoost)

Algoritmul **XGBoost**, recunoscut pentru eficiența sa în modelarea relațiilor complexe, a fost implementat pentru a estima prețurile apartamentelor, utilizând arbori de decizie și tehnici de boosting. Antrenarea modelului a implicat un proces riguros de **optimizare a hiperparametrilor**, desfășurat exclusiv pe setul de antrenament validat prin **5-fold cross-validation**.

Setul extins de hiperparametri a inclus:

- **numărul de arbori** (nrounds),
- **adâncimea arborilor** (max_depth),
- **rata de învățare** (eta),
- **coeficienți de regularizare** (gamma, min_child_weight),
- **procentul de caracteristici selectate** (colsample_bytree).

Combinarea acestor parametri a generat peste **5.000 de configurații testate**, iar performanța fiecărui model a fost evaluată pe baza **RMSE, R^2 și MAE**. Cea mai bună configurație a inclus:

- **eta = 0.05,**
- **max_depth = 2,**
- **gamma = 10,**
- **colsample_bytree = 0.75,**
- **min_child_weight = 1,**
- **750 iterații,**
- și a condus la un **RMSE de 16.706,96, $R^2 = 82,3\%$ și MAE = 11.551,2.**

Modelul a demonstrat o **excelentă capacitate de generalizare**, echilibrând corect complexitatea cu acuratețea, datorită mecanismelor avansate de regularizare și ajustare iterativă a erorilor. În plus,

analiza importanței variabilelor prin funcția `xgb.importance` a evidențiat factorii cei mai influenți în determinarea prețului final.

Prin performanțele sale, **XGBoost s-a dovedit a fi cel mai precis model** testat în cadrul acestei cercetări, depășind metodele anterioare printr-o abordare robustă, scalabilă și eficientă în tratarea relațiilor non-liniare.

4.5 Rezultate obținute

Studiul de caz a aplicat cinci metode de învățare automată pentru estimarea prețurilor apartamentelor din Cluj-Napoca, utilizând aceeași metodă de validare (5-fold cross-validation) și același set de date. Compararea performanței modelelor s-a realizat pe baza valorilor RMSE.

4.5.1 Analiza comparativă a metodelor aplicate

Modelele testate au inclus regresia liniară, arborii de decizie, pădurile aleatorii, rețelele neuronale artificiale (ANN) și XGBoost. Valorile RMSE au arătat o diferențiere clară a performanței:

- **XGBoost:** 16.706,96
- **ANN:** 16.940,07
- **Random Forest:** 17.533,62
- **Regresie liniară:** 17.780,00
- **Arbori de decizie:** 18.052,18

Rezultatele validează ipoteza de cercetare: metodele complexe precum XGBoost și ANN oferă o precizie superioară în predicție, reușind să capteze mai bine relațiile dintre variabilele explicative și prețul apartamentelor.

4.5.2 Performanța modelului XGBoost în estimarea prețurilor

Modelul XGBoost, antrenat pe 80% din date și testat pe 20%, a obținut un **RMSE de 16.706,96** și un **MAE de 11.551,20**. Diferențele dintre valorile reale și cele estimate sunt moderate, iar un exemplu relevant a fost o predicție cu doar 44 Euro abatere față de prețul real.

Graficul aferent a evidențiat o corelație ridicată între valorile estimate și cele reale, confirmând capacitatea modelului de a surprinde tendințele generale ale pieței imobiliare. Totuși, în unele cazuri extreme, au fost observate deviații, semnalând zone de îmbunătățire.

4.5.3 Evaluarea finală a modelului XGBoost pe setul de testare

În testarea finală pe un set complet separat (20% din date), XGBoost a obținut:

- **RMSE = 20.457,68**
- **Interval de încredere RMSE (95%) = 17.260,76 – 23.654,61**

Aceste rezultate validează robustețea modelului, dar evidențiază și limitele impuse de factori precum: mărimea setului de date, dezechilibrul geografic sau absența unor variabile explicative suplimentare (ex. factori economici, subiectivi).

4.6 Discuții, interpretări, limitări și concluzii finale

Subcapitolul 4.6 sintetizează implicațiile teoretice și practice ale cercetării, oferind o interpretare critică a rezultatelor, identificând limitările studiului și formulând concluziile finale privind performanța metodelor de învățare automată aplicate în evaluarea prețurilor imobiliare.

4.6.1 Discuții și limitări

Analiza s-a bazat pe un set de date colectate manual din Cluj-Napoca, fiind testate cinci modele predictive, de la tehnici tradiționale (regresie liniară, arbori de decizie) până la modele avansate (Random Forest, ANN, XGBoost). Rezultatele au confirmat superioritatea metodelor moderne, în special a XGBoost, însă performanța a fost condiționată de factori precum dimensiunea redusă a eșantionului, distribuția inegală a datelor pe cartiere și dezechilibrul unor variabile categoriale.

Limitările identificate includ:

- dimensiunea restrânsă a bazei de date, influențată de colectarea manuală;
- dezechilibre geografice și categoriale în distribuția apartamentelor;
- subreprezentarea unor subcategorii, care reduce capacitatea de generalizare a modelelor.

Se recomandă, pentru viitor, automatizarea procesului de colectare a datelor și extinderea acestuia pentru a asigura echilibrul și acuratețea estimărilor.

4.6.2 Concluzii finale

Compararea modelelor a evidențiat performanțele superioare ale XGBoost, care a obținut cel mai mic RMSE (16.706,96) și o explicație statistică robustă ($R^2 = 0,823$). Tabelul comparativ al variabilelor semnificative a arătat că anumite atribute, precum suprafața, numărul de camere, etajul sau zona, sunt recurente în toate modelele, validându-le importanța în procesul de evaluare imobiliară.

De asemenea, metodele avansate au demonstrat o capacitate crescută de a capta relații complexe între variabile, identificând factori relevanți care nu sunt detectați de modelele tradiționale. Utilizarea în paralel a mai multor algoritmi poate oferi predicții mai robuste și cuprinzătoare.

În concluzie, studiul validează utilitatea învățării automate în estimarea valorii proprietăților rezidențiale, subliniind importanța selecției variabilelor, optimizării hiperparametrilor și extinderii bazelor de date. Cercetarea oferă o bază solidă pentru dezvoltarea unor sisteme automate de evaluare, adaptate pieței imobiliare din România.

Concluzii și direcții viitoare de cercetare

Lucrarea a demonstrat că metodele de învățare automată pot îmbunătăți semnificativ procesul de evaluare a proprietăților imobiliare, oferind un cadru predictiv mai rapid, mai obiectiv și mai adaptabil decât abordările tradiționale. Plecând de la fundamentele teoretice și metodologice ale evaluării (Cap. 1), studiul a susținut integrarea AVM-urilor ca soluție inovatoare în contextul actual al pieței imobiliare.

Analiza literaturii (Cap. 2) a confirmat creșterea interesului academic pentru AVM-uri și a evidențiat atât avantajele acestora în piețele mature, cât și provocările de implementare în contexte locale, precum România. Capitolul 3 a aprofundat aplicabilitatea metodelor moderne – regresie, arbori de decizie, păduri aleatorii, rețele neuronale și XGBoost – evidențiind criteriile de selecție a modelului optim în funcție de scop și de calitatea datelor disponibile.

Studiul de caz din Cluj-Napoca (Cap. 4) a validat ipoteza conform căreia modelele complexe, în special XGBoost, oferă o acuratețe superioară în estimarea prețurilor rezidențiale. Modelul XGBoost a obținut cel mai mic RMSE și MAE, confirmându-și superioritatea în raport cu celelalte metode analizate. Cu toate acestea, limitările precum dimensiunea redusă a setului de date și dezechilibrele variabilelor categorice impun prudență în generalizarea rezultatelor.

Concluziile susțin utilizarea unui cadru hibrid: combinarea inteligenței artificiale cu expertiza umană în procesul de evaluare. Direcțiile de cercetare propuse includ extinderea pe seturi de date regionale și temporale variate, integrarea de factori macroeconomici și explorarea unor tehnologii emergente precum IoT sau blockchain. Totodată, viitoarele cercetări ar trebui să vizeze interpretabilitatea modelelor și influența sustenabilității asupra valorilor imobiliare.

În ansamblu, lucrarea oferă o contribuție relevantă atât teoretic, cât și practic, confirmând potențialul real al învățării automate în redefinirea standardelor de evaluare imobiliară.