PROGNOZAREA ADMITERII UNIVERSITARE BAZATĂ PE MODELE STATISTICE ȘI INTELIGENȚĂ ARTIFICIALĂ

UNIVERSITARY ENROLLMENT FORECASTING THROUGH STATISTICAL AND AI-BASED MODELING

CZU: 004.89:378 **https://doi.org/10.59295/spd2025e.104**

Dan LEFTER
ORCID: 0009-0008-3240-6642
E-mail: dan.lefter@usm.md
Universitatea de Stat din Moldova, Chișinău, Republica Moldova

Maria CAPCELEA
ORCID: 0009-0009-3321-3946
E-mail: maria.capcelea@usm.md
Universitatea de Stat din Moldova, Chișinău, Republica Moldova

Summary.*This paper presents a predictive analysis system designed to estimate the future evolution of student enrollment in Moldovan universities. Using official data on natality (1980–2023), student enrollment, and graduate statistics (2005–2023), the study employs a multi-model approach based on statistical indicators and machine learning forecasting methods. Models such as linear regression, polynomial regressions (degrees 2 and 3), exponential growth, and autoregressive (AR) analysis were implemented and compared. The models are integrated into an interactive Streamlit web application, allowing stakeholders to select forecasting horizons and visualize the future trends of university enrollment by specialty. Results show differentiated evolutions among specialities and highlight the importance of demographic trends in educational planning [1-2].*

Keywords: *autoregression, educational forecasting, enrollment prediction, machine learning, student statistics.*

INTRODUCERE

Transformările demografice, sociale și economice din ultimele decenii au generat schimbări semnificative în cererea pentru educație superioară în Republica Moldova. În acest context, previzionarea evoluției numărului de studenți devine o necesitate esențială pentru instituțiile de învățământ superior și pentru factorii decizionali de la nivel național, în concordanță cu tendințele globale de participare la învățământul superior raportate de World Bank [3].

Obiectivul principal al lucrării este de a elabora un sistem de analiză predictivă a numărului de studenți în funcție de evoluția natalității și de tendințele în învățământul superior. Acest sistem se bazează pe metode statistice clasice, completate cu tehnici mo-

derne de inteligență artificială, și este integrat într-o platformă interactivă destinată utilizatorilor din domeniul educației și cercetării [4-5].

Lucrarea este structurată astfel: după prezentarea cadrului metodologic și a seturilor de date utilizate, sunt analizate modelele de predicție și rezultatele obținute. În final, se oferă concluzii, limitări și direcții viitoare de cercetare și implementare.

MATERIAL ȘI METODĂ

Studiul utilizează date oficiale provenite din următoarele surse:

- **Natalitatea anuală (1980–2023)** – date statistice naționale publicate de Biroul Național de Statistică [1];

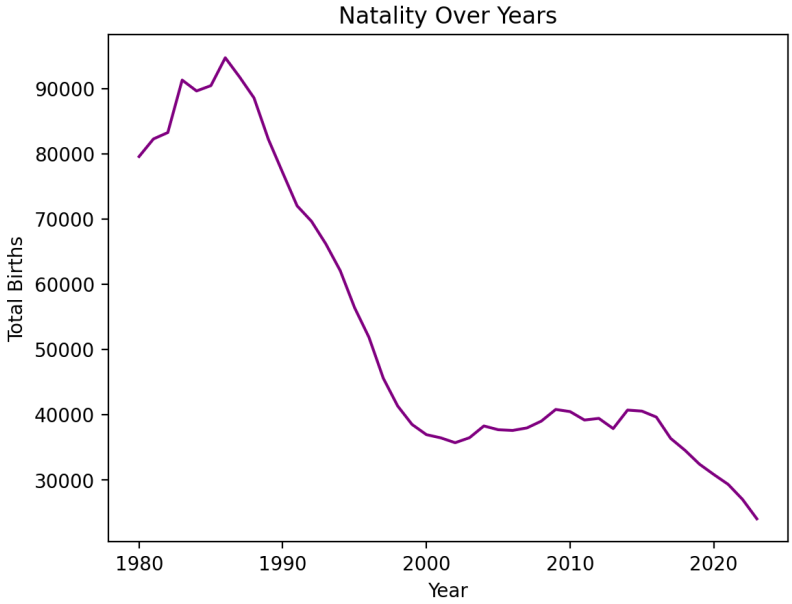


Figura 1. Natalitatea pentru perioada 1980 - 2023

- **Numărul de studenți înscriși și absolvenți (2005–2023)** – statistici educaționale oficiale furnizate de Ministerul Educației și Cercetării [2];

Datele au fost prelucrate folosind biblioteci Python precum *pandas*, *numpy*, *scikit-learn*, *matplotlib*, *statsmodels* și *streamlit* [5].

Pașii metodologici au inclus:

1. **Curățarea și transformarea datelor**:
   - Eliminarea valorilor lipsă;
   - Corelarea anului nașterii cu anul posibil de admitere (presupunând vârsta de 18 ani);
   - Alinierea seriilor de timp pe aceeași axă cronologică.

2. **Calculul indicatorilor descriptivi**:
   - Media, mediana, variația standard, coeficientul de variație, pantele trendului anual;
   - Selecția specialităților cu cel mai mare și cel mai mic număr de înscriși pe an [6].
3. **Modele aplicate:**
   - **Regresie liniară**: bună pentru tendințe uniforme, aplicată pentru serii stabile (ex. *Informatică*);
   - **Regresie polinomială (gradul 2 și 3)**: captarea curbelor cu oscilații, util pentru serii precum *Arte* sau *Economie*;
   - **Model exponențial**: folosit pentru serii cu creștere accelerată sau scădere abruptă;
   - **Autoregresie (AR)**: folosit în special pentru serii sezoniere și ciclice precum *Pedagogie*, oferind precizie ridicată [4].
4. **Evaluarea modelelor**:
   - R^2, MSE și MAE au fost calculate pentru fiecare model și specialitate;
   - Intervalele de încredere 95% au fost reprezentate grafic [5].
5. **Integrarea aplicației în Streamlit**:
   - Utilizatorii pot selecta specialitatea, modelul de predicție și perioada de prognoză (până în 2035);
   - Graficele interactive permit compararea valorilor reale și estimate.

REZULTATE ȘI DISCUȚII

Rezultatele obținute indică o evoluție diferită în funcție de domeniul de studiu:

- *Ştiințe ale educației*: compararea modelelor indică o superioritate clară a modelului autoregresiv (AR), care a atins un coeficient R^2 de 0.9819, față de doar 0.5685 în cazul regresiei exponențiale. De asemenea, intervalul de încredere de ±993.96 în AR confirmă o stabilitate ridicată a predicției, comparativ cu ±4694.64 în modelul exponențial. Declinul estimat este de -4.37% până în 2033 în AR, față de -15.68% în modelul exponențial. Aceste diferențe evidențiază necesitatea utilizării unor modele robuste în domeniile sensibile precum educația.
- *Arte* înregistrează o tendință relativ stabilă, cu variații moderate și o ușoară creștere estimată până în 2030;
- *Agronomie* și *Ştiințe exacte* prezintă o scădere continuă, corelată cu natalitatea redusă după anii 2000 [2];
- *Informatică* prezintă o creștere ușoară, dar instabilă, captată mai bine de regresia polinomială de gradul 3;
- Pentru *Pedagogie*, modelul AR a prezentat cel mai mic MAE (~48 studenți/an) și un R^2 de peste 0.92, indicând un model optim.

Tabel 1. Comparație între metodele de regresie

Model	Utilitate	Avantaje	Limitări
Liniară	Serii stabile, tendințe simple	Ușor de interpretat	Sensibil la outliers
Polinomială	Oscilații, sezonalitate slabă	Flexibilă, potrivită pentru Arte	Risc de overfitting
Exponențială	Scădere/creștere rapidă	Captare rapidă a declinului	Sensibilă la fluctuații
Autoregresiv	Serii ciclice sau sezoniere	Precizie ridicată (ex: AR R^2=0.98)	Necesită istoric lung

Mai mult, dacă analizăm rata de creștere sau descreștere pentru fiecare model, observăm că predicțiile variază semnificativ. Spre exemplu, în *Științele educației*, modelul exponențial estimează o descreștere de -15.68%, ceea ce implică un declin abrupt, pe când modelul AR anticipează o descreștere moderată de -4.37%, într-un interval de încredere mult mai îngust. Acest contrast are implicații directe în planificarea resurselor și a finanțării universitare.

Natalitatea din Republica Moldova a suferit o scădere accentuată după anii 2000, determinată de factori economici, sociali și migrațională [1]. Acest declin demografic se reflectă direct în scăderea populației tinere eligibile pentru învățământul superior. Pe fondul acestei tranziții, instituțiile educaționale se confruntă cu necesitatea de a adapta oferta academică, infrastructura și strategiile de recrutare.

Corelarea natalității cu numărul de studenți indică o întârziere de aproximativ 18 ani între scăderea numărului de nou-născuți și reducerea cohortelor de admitere la facultate. În acest sens, utilizarea predicțiilor pe baza datelor demografice devine crucială pentru politici publice coerente, estimarea bugetelor instituționale și dimensionarea locurilor bugetate [2].

Utilizarea corectă a modelelor predictive permite instituțiilor să adopte măsuri proactive. Spre exemplu:

- **Scăderile abrupte prognozate** (ex: -15.68% în modelul exponențial pentru Științele Educației) pot avertiza asupra reducerii necesare a capacităților instituționale – săli, cadre, locuri bugetate;
- **Modelul autoregresiv**, care prezice o scădere mai lentă (-4.37%), permite un **management gradual** al resurselor;
- **Intervalele de încredere mai înguste**, precum ±993.96 în modelul AR, sporesc fiabilitatea deciziilor bazate pe aceste prognoze, în comparație cu cele largi (±4694.64) care induc incertitudine.

Aceste interpretări pot fi extrapolate la alte specialități. În cazul Informaticii, creșterea neregulată dar general pozitivă poate sugera oportunitatea extinderii capacităților în acest domeniu. În contrast, domenii precum Agronomia sau Științele exacte, cu o tendință descendentă constantă, necesită o regândire a politicilor de promovare și atragere a studenților.

Pentru extinderea acestei cercetări, sunt planificate următoarele:

- **Integrarea datelor socio-economice** (venitul mediu pe gospodărie, rata șomajului) în modelele predictive, pentru a analiza cum influențează acestea decizia de a urma studii superioare;
- **Includerea unui model LSTM (Long Short-Term Memory)** pentru a compara performanța rețelelor neuronale cu modelele statistice clasice;
- **Publicarea aplicației ca serviciu educațional web** cu acces extins pentru instituții, prin integrarea de API-uri cu surse oficiale (BNS, MEC);
- **Studii de impact**: estimarea pierderilor financiare instituționale dacă predicțiile nu sunt luate în considerare (analiză de risc).

CONCLUZIE

Studiul de față demonstrează aplicabilitatea metodelor statistice și algoritmice în previzionarea evoluției învățământului superior din Republica Moldova. Modelele AR s-au dovedit superioare în stabilitate și precizie, în special pentru domeniile cu sezonalitate pronunțată. Legătura clară între natalitate și înmatriculare validează ipoteza că previziunile educaționale trebuie corelate cu dinamica populației.

În concluzie, predicțiile pe baza datelor reale nu doar că pot îmbunătăți eficiența instituțională, dar pot susține și formularea politicilor educaționale adaptate realităților demografice și socio-economice.

REFERINȚE:

1. BIROUL NAȚIONAL DE STATISTICĂ AL REPUBLICII MOLDOVA. Date statistice despre populație și educație. Disponibil: https://statistica.gov.md [Accesat: 08.02.2025].
2. MINISTERUL EDUCAȚIEI ȘI CERCETĂRII. Anuare statistice educaționale 2005–2023. Disponibil: https://mecc.gov.md [Accesat: 28.04.2025].
3. WORLD BANK. Higher Education Participation Rates, 2023. Disponibil: https://data.worldbank.org [Accesat: 08.07.2025].
4. HYNDMAN, R.J., ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. Monash University, 2021. Disponibil: https://otexts.com/fpp3/ [Accesat: 18.05.2025].
5. STREAMLIT. Documentation. Disponibil: https://docs.streamlit.io [Accesat: 05.03.2025].
6. SCIKIT-LEARN. Machine Learning in Python. Disponibil: https://scikit-learn.org [Accesat: 02.03.2025].

N.B. *Această lucrare reprezintă rezultatul activității de cercetare desfășurate în cadrul proiectului „011302 Metode analitice și numerice pentru rezolvarea problemelor decizionale dinamice stochastice”.*