THE USE OF DATA ANALYSIS TO ANALYZE EDUCATION TRENDS IN AZERBAIJAN

Jabbarova K.,
PhD, *associate professor of Azerbaijan State Oil and Industry University*
Ismayilov H.
master's student, Azerbaijan State Oil and Industry University
https://doi.org/10.5281/zenodo.15536101

Abstract

In recent years, the importance of understanding educational trends has significantly increased, especially in developing countries like Azerbaijan. This study focuses on using data analysis methods to examine and predict changes in the field of education. The aim is to analyze educational datasets and identify key indicators that reflect long-term trends, such as student enrollment rates, graduation statistics, and public spending on education. Using Python-based tools and time series forecasting models, this research explores how these indicators change over time and what they suggest about the future of education in Azerbaijan. The study also applies statistical models such as linear regression and autoregressive models to perform trend analysis and generate forecasts. The findings of this study can help policymakers, educators, and institutions make data-driven decisions to improve the quality and accessibility of education.

Keywords: Education Trends, Data Analysis, Time Series Forecasting, ARIMA, Autoregressive Models, Stationarity

Introduction

In today's rapidly evolving global landscape, education has become a cornerstone for national development, especially in emerging economies such as Azerbaijan. Understanding the dynamics and trends within the education sector is crucial for shaping policy, allocating resources effectively, and forecasting future needs [1].

In recent years, the role of data analysis in education has grown significantly, enabling institutions and governments to make informed, evidence-based decisions [2]. Time series forecasting models such as Autoregressive (AR), Integrated (I), and Moving Average (MA) components have been successfully applied in multiple domains, including education and economics [3].

This study focuses on analyzing long-term education trends in Azerbaijan using data-driven techniques. By leveraging historical datasets on higher education enrollment, we aim to uncover key patterns and predict future student admission numbers. Through the application of time series forecasting methods—namely, autoregressive (AR), differenced autoregressive (stationary AR), and ARIMA models—we intend to evaluate the accuracy and reliability of various forecasting approaches.

By incorporating programming tools such as Python and libraries including pandas, statsmodels, and matplotlib, the study demonstrates how modern data science methods can be used not only to visualize educational trends but also to generate actionable insights for policymakers and academic institutions [4].

Dataset Description

The dataset used in this study comprises the number of students admitted to higher education institutions in Azerbaijan, covering the academic years from 2000/2001 to 2023/2024. The data was obtained from the official website of the State Statistical Committee of the Republic of Azerbaijan (stat.gov.az) and includes annual admission figures for both public and private universities [5]. However, due to missing data for the academic years 2001/2002 through 2004/2005, linear interpolation was employed to estimate the absent values. This preprocessing step was crucial to ensure continuity in the time series, a requirement for building accurate and consistent forecasting models.

The final dataset consists of 24 academic years, structured in two columns: **Year**, which records the academic year in the format YYYY/YYYY+1, and **Admissions**, which indicates the number of students admitted to bachelor's programs. The cleaned and interpolated dataset was saved in an Excel file and served as the foundational input for all forecasting models used in this research. A sample view of the dataset is presented below.

Year	Admissions
2000/2001	26,403
2001/2002	27,575
2002/2003	28,747
…	…
2023/2024	52,880

All values are rounded to the nearest whole number, and the data structure is suitable for applying time series forecasting models using Python.

Methodology

To forecast future student admissions in Azerbaijan, we utilized three distinct time series models. First, we applied an Autoregressive (AR) model to the original non-stationary data. Next, we implemented another AR model, this time using stationary data obtained through differencing. Finally, we employed the ARIMA (Autoregressive Integrated Moving Average) model, which combines both autoregressive and moving average components along with differencing to handle non-stationarity [1][2].

All models were implemented in Python using pandas for data manipulation, statsmodels for statistical modeling, and matplotlib for visualization [4]. Forecasts were generated for two future academic years: **2024/2025** and **2025/2026**.

AR Model on Non-Stationary Data: The autoregressive model (AR) uses past values of a time series to predict future values. The non-stationary version of this model was applied directly to the raw admissions data, without transforming it to achieve stationarity. A lag of 3 was selected based on preliminary testing, which means the model uses the previous three years of data to predict the next value [2]. Model equation:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \epsilon_{ta}$$

AR Model on Stationary Data (Differencing): Since the original admissions data exhibited a clear upward trend, we applied first-order differencing to transform the series into a stationary form, which is a common requirement for many time series models [1][3]. After differencing, the same AR(3) model was applied to the transformed data. The different forecasts were then added cumulatively to the last known admission value to obtain the predicted future values.

ARIMA Model: The ARIMA model integrates three key components to effectively model time series data. The **Autoregression (AR)** component utilizes lagged values of the variable to capture temporal dependencies. The **Integration (I)** component applies differencing to the data to achieve stationarity, which is essential for accurate forecasting. Lastly, the **Moving Average (MA)** component incorporates past forecast errors to refine predictions and improve model accuracy [3].

For this study, we used an **ARIMA(1,1,0)** configuration based on data behavior and model performance. This setting includes first-order differencing and one autoregressive term. Each model was trained separately and saved using joblib, and a unified forecast_and_plot() function was used to load the models, generate forecasts, and visualize results. All source code used in this study, including data preprocessing, model training, and visualization, is publicly available on GitHub:

https://github.com/HesenIsmayilov44/education-trend-analysis-azerbaijan.git *(Accessed on May 2025)*

Results And Forecasting

Each of the three forecasting models was trained using the interpolated dataset covering the years 2000/2001 to 2023/2024. The goal was to predict the number of students expected to be admitted to higher education in Azerbaijan for the academic years **2024/2025** and **2025/2026**.

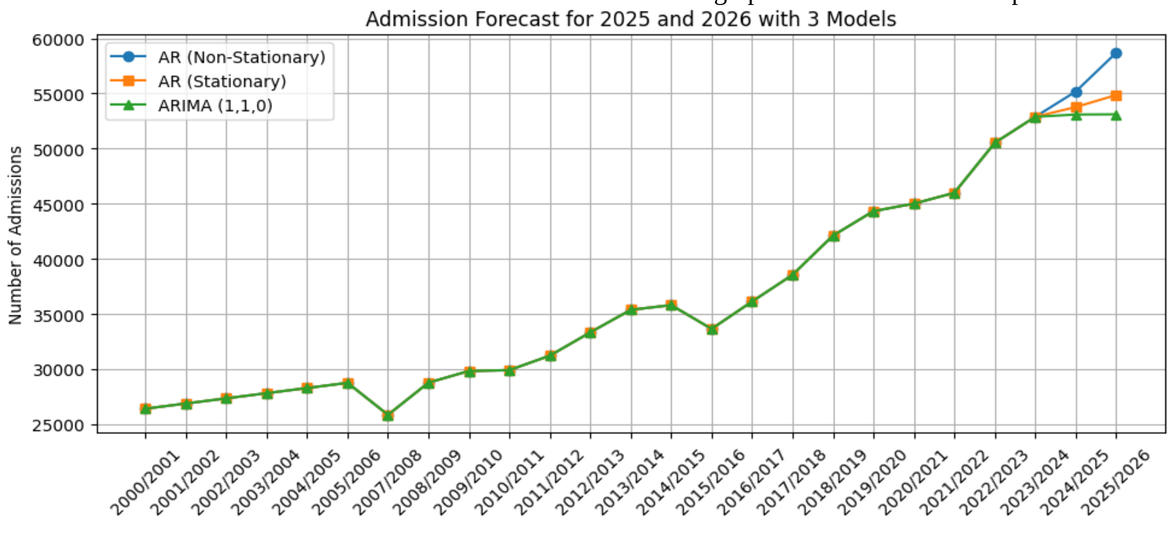
Forecasted Values

Model	2024/2025	2025/2026
AR	55,169	58,687
AR (Stationary)	53,765	54,853
ARIMA	53,087	53,106

The AR (non-stationary) model generated a more aggressive upward forecast, extending the recent increasing trend in student admissions. In contrast, the AR model applied to stationary (differenced) data produced a more conservative projection, stabilizing the growth rate. The ARIMA model indicated a modest increase, likely due to the influence of its moving average and differencing components, which help moderate sharp fluctuations in the data.

Visualization

The following chart presents all three models on the same graph for a clear visual comparison:



In the comparative visualization, the x-axis represents academic years, while the y-axis displays the number of admitted students. For each model, two forecast years are appended at the end of the original dataset to illustrate the projected values. This side-by-side comparison highlights both the trend and the range of predictions generated by each approach. Despite differences in methodology, all models consistently indicate a continued growth in student admissions.

Discussion

The results from the three time series models provide valuable insights into the possible future of student admissions in Azerbaijan. Each model captures different aspects of the dataset, and together, they offer a comprehensive understanding of trend behavior and model reliability.

Model Comparison

AR (Non-Stationary) This model directly uses the raw data without transformation. It performs well in capturing existing upward trends but may overfit in the presence of structural breaks or long-term noise. Its forecasts suggest a sharp increase in future admissions, which may be realistic if growth continues as in the past, but it lacks resistance to anomalies.

AR (Stationary) By differencing the data first, this model removes trend and seasonality, allowing for more stable predictions. Its results are more conservative and arguably more reliable over longer horizons. This model is suitable when the assumption is that growth will slow down or stabilize over time.

ARIMA (1,1,0) Combining autoregression and differencing, this model finds a balance between stability and adaptability. Its forecasts indicate moderate, steady growth. ARIMA is particularly suitable when the data shows both trend and short-term fluctuation, as it captures both long-term direction and local variations.

Interpretation in Educational Context

From a policy perspective, even the most conservative model—the ARIMA model—predicts a continued increase in higher education admissions. This trend implies a growing demand for university infrastructure, faculty, and academic programs, highlighting the need for improved funding and strategic planning within the education sector. It also suggests potential pressure on student services, dormitory capacities, and the employment market following graduation. Furthermore, the models offer a solid foundation for expanding the analysis by incorporating external variables such as population size, economic indicators, or government education policies, which could provide a more comprehensive understanding of the factors influencing student admissions.

Conclusion

This study explored the application of time series forecasting techniques to predict future trends in higher education admissions in Azerbaijan. Using historical data from 2000/2001 to 2023/2024, we implemented and compared three models: AR (non-stationary), AR (stationary), and ARIMA. The findings from all three models consistently indicated continued growth in the number of students entering higher education over the next two academic years. While the AR model projected a more aggressive increase, the ARIMA and stationary AR models provided more moderate and conservative forecasts.

From a strategic standpoint, the results of this study can assist educational policymakers in anticipating future demand for university seats, enabling more efficient allocation of educational resources. The insights also support informed decision-making regarding admissions quotas and academic staffing, ensuring that institutions are better prepared to accommodate projected growth in student numbers. In addition, the study demonstrated the importance of data preprocessing (e.g., interpolation and differencing) and model selection in ensuring reliable forecasts.

Future Work

Future enhancements to this research may involve incorporating exogenous variables such as population growth, migration trends, or government funding levels through the use of ARIMAX or machine learning models. Expanding the forecast horizon beyond 2026 could provide a longer-term perspective on admissions trends. Additionally, applying ensemble or hybrid modeling approaches may improve forecast accuracy. Ultimately, this study underscores the value of data-driven analysis in supporting forward-looking, evidence-based decision-making within the education sector.

References:

1. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.). OTexts. https://otexts.com/fpp2/

2. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). Wiley.

3. Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer.

4. Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. Proceedings of the 9th Python in Science Conference. https://www.statsmodels.org/

5. State Statistical Committee of the Republic of Azerbaijan. (2025). Education statistics. Retrieved from https://www.stat.gov.az/source/education/