Time series forecasting of internal student migration using ML-ARIMA model

*N. Jayenta Meitei[1], Md. Baharuddin Shah[2]
[1]Assistant Professor, Department of Geography, Y.K. College, Wangjing, Manipur
[2]Md. Baharuddin Shah, Professor, Department of Geography, G.P. Women's College, Imphal, Manipur

*Corresponding Author: N. Jayenta Meitei
Email: jayenta.kak@gmail.com

Abstract

This study employs a machine learning-based predictive approach to assess the trend of internal migration of students from the Northeast region of India between 2011 and 2031. The study uses Census of India migration data from 1981 to 2011 to use the ARIMA model as part of a predictive inferential research approach. This study aims to close the data gap on internal student migration from Northeast India, where there has been a relatively high intensity of educational out-migration. The research predicts that the amount of this migratory stream will keep on rising. The ARIMA model shows a consistent rise in student out-migration from the region. The results point to significant policy ramifications for resolving the region's educational infrastructure shortcomings and lessening the strains imposed by migrants on big urban educational hubs.

Keywords: Machine learning, student/educational migration, ARIMA model, Northeast India

Introduction

In numerous domains, machine learning (ML) and predictive analytics have become transformative instruments, facilitating data-informed decisions that enhance productivity, accuracy and strategic planning. Machine learning has established itself as a groundbreaking analytical approach for uncovering concealed patterns, anticipating both institutional and human behaviours, and refining decision-making across diverse fields such as migration studies, education, healthcare and organizational management. The integration of artificial intelligence and machine learning signifies a fundamental shift towards data-driven governance and strategic planning encompassing areas from population mobility forecasting to healthcare resource management and the prediction of educational outcomes (Hussain, 2021; Avinash, G. et al., 2025). India's Northeast region (NER), which has been a major source of out-migration for employment and education in the interstate migrant flows over the past few decades (Lyndem & De, 2004; Shimray & Usha, 2009; Chyrmang, 2011; McDuie-Ra, 2012; Lusome & Bhagat, 2020), is an important research area to apply ML-based predictive analysis. The Northeastern states of India encompassing Assam, Arunachal Pradesh, Sikkim, Meghalaya, Nagaland, Manipur, Mizoram and Tripura exhibit a notably higher proportion of student mobility compared to the national average (Mistri & Sardar, 2022). Due to the postponement of the Census of India since 2011, there is a paucity of updated data concerning migration trends and patterns with the exception of sample survey data. Consequently, this study aims to address the existing knowledge deficit regarding the contemporary trends of internal student out-migration from India's Northeastern states during the period spanning 2011 to 2031.

Literature Review

The ARIMA model is a widely used statistical tool in predictive research especially for analyzing and predicting time series data. This model introduced by Box and Jenkins in 1970 captures the relationships and trends in a dataset by combining autoregressive (AR), differencing (I), and moving average (MA) components. It's particularly effective for short-term forecasting when the data shows non-stationarity that can be stabilized by differencing (Box & Jenkins, 1970; Hyndman & Athanasopoulos, 2018). Since ARIMA models can identify underlying temporal patterns without needing outside predictors, they are used in various fields including economics, population studies, climatology and migration forecasting (Wei, 2006). Model selection and predictive accuracy are improved by recent developments that combine ARIMA with machine learning methods, such as auto-ARIMA algorithms or hybrid ARIMA–ANN (Zhang, 2003; Makridakis et al., 2018). Educational migration especially from peripheral regions like Northeast India exhibits complex temporal dynamics influenced by socioeconomic, institutional and policy factors (Nongkynrih, 2020; McDuie-Ra, 2013). New developments combine ARIMA with machine learning methods such as hybrid ARIMA–ANN (Artificial Neural Network) or auto-ARIMA to increase accuracy, manage nonlinearity and automate model selection (Zhang, 2003; Makridakis et al., 2018). These ML models enable more accurate predictions of student departures in educational migration studies, which help policymakers anticipate the demand for higher education, regional skill migration and infrastructure requirements (Das & Saikia, 2019).

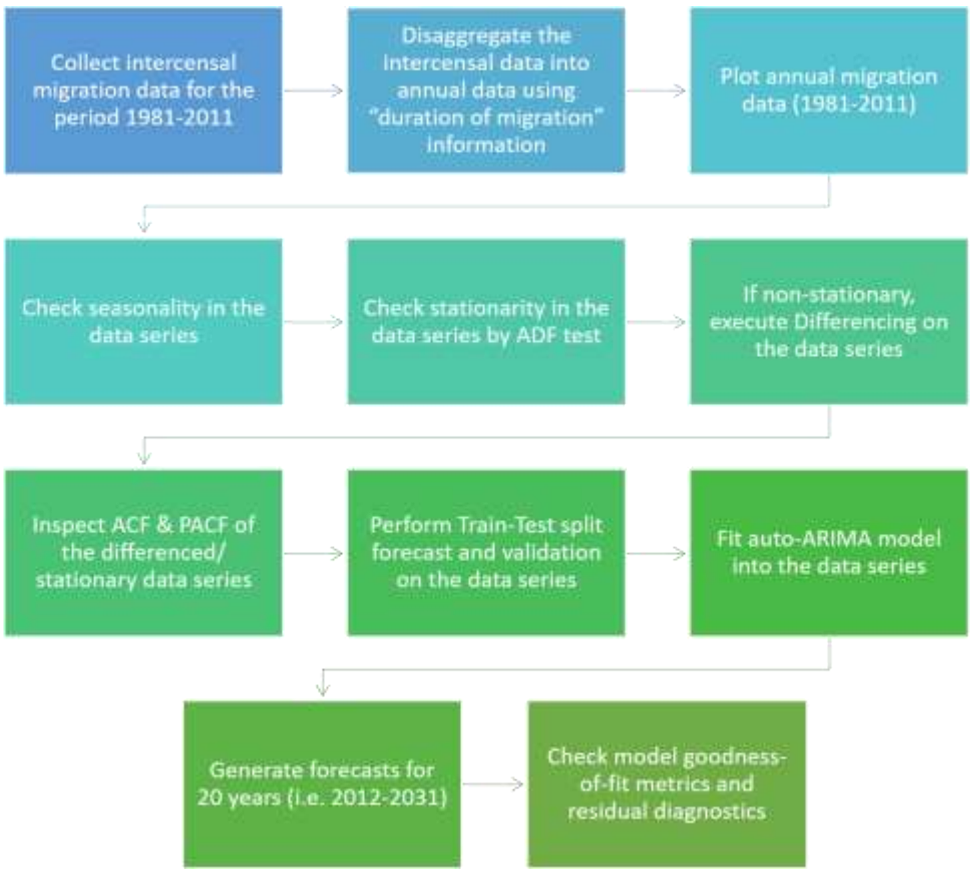
The way complex real-world phenomena are modelled, interpreted and predicted in research has been completely transformed by the incorporation of machine learning into predictive analytics. In machine learning-based predictive analysis, algorithms that identify patterns in data are used to forecast future events with little assistance from humans (Jordan & Mitchell, 2015). Traditional statistical methods, like linear regression or ARIMA models depend on clear presumptions regarding the stationarity and distribution of data. By contrast, machine learning (ML) algorithms like Random Forests, Support Vector Machines (SVM) and Neural Networks, are particularly effective in high-dimensional and dynamic data contexts because they can capture complex nonlinear relationships and interactions among variables (Murphy, 2012; Hastie et al., 2009).These models are increasingly being applied in domains ranging from economics and health to education and environmental forecasting, where predictive precision and adaptability are crucial (Padmaja et al. 2024; Vatti et al., 2024). Over time, hybrid models that integrate traditional time-series approaches like ARIMA with machine learning have gained prominence. These hybrid systems combine the ability of machine learning algorithms to learn adaptively with the interpretability of statistical models (Zhang, 2003; Makridakis et al., 2018). For example, researchers can model both the linear components (represented by ARIMA) and nonlinear residuals (managed by neural networks) using hybrid ARIMA–ANN models, which improves forecasting accuracy (Zhang, 2003). By further automating model selection, hyperparameter tuning, and feature optimization, the rise of auto-ML frameworks and auto-ARIMA algorithms has decreased human bias in predictive analytics (Hyndman & Athanasopoulos, 2018). More reliable and adaptable predictions across disciplines are made possible by the methodological advancement in forecasting sciences represented by this combination of the classical and data-driven modelling paradigms.

In social science and migration research, where datasets frequently display abnormalities, missing values and nonlinear behavioural patterns, machine learning-based predictive analysis has become especially pertinent in recent years (Amini et al., 2018). By utilizing a variety of socioeconomic indices, machine learning approaches have been utilized to forecast trends in student migration, labour mobility and demographic changes (Das & Saikia, 2019; Yang et al., 2020). These applications show how evidence-based policymaking can be informed by predictive analytics particularly in areas where data is few or situations are changing quickly. Researchers can improve model dependability and offer more nuanced insights into educational flows, policy impacts, and temporal and spatial migration dynamics by integrating machine learning with conventional forecasting techniques (Bishop, 2006).

Methodology

The current study uses a machine learning approach in conjunction with a predictive inferential research strategy. Figure 1 shows the specific steps involved in the research. This study uses an Auto-regressive Integrated Moving Average (ARIMA) model to accomplish its goal. Three fundamental tasks—"split," "train," and "test"—are involved in ML-based predictive analysis and must be carried out in a forecasting or prediction model. This is accomplished in the current work by dividing the data series into two segments, training the model to identify past trends from one segment, and evaluating the forecast performance using the remaining segment. In order to make predictions for the future, the best forecast model is built using model fitness measures. The Python package version 3.13 was used for all statistical analysis and modelling in this study. *NumPy, Pandas, Itertools, Statsmodels, Scikit-learn,* and *Matplotlib* are among the Python libraries utilized in this investigation.

Figure-1:
Steps of ML-ARIMA modelling & forecasting of student out-migration



Data Sources

The stock of educational out-migrants from India's Northeastern states between 1981 and 2011 is used in this analysis. The D-3 tables from the Indian Census of 1981, 1991, 2001 and 2011 are the source of this information. Interstate educational out-migrants (migration for the purpose of education) whose last usual place of residence is any of the eight Northeast states of India are taken out of the Census tables for analysis based on the information provided on "reasons of migration" and "last usual place of residence." The years 1981–2011 are taken into consideration because the 2011 Census was the last one carried out in the nation, and the Census started gathering data on the "education" reason of migration in 1981.

Table-1: Stock of educational out-migrants from the Northeast states of India

Year of census	Intercensal migrants	Total stock of migrants
1981[#]	--	18,000
1991	17,884	26,327
2001	26,659	30,173
2011	32,737	40,478

Data source: Computed from D-3 tables of Census of India 1981, 1991, 2001, 2011.
[#]Census was not conducted in the state of Assam in Northeast India due to socio-political disturbances in 1981. Hence, the total stock of migrants is estimated by interpolation.

This study gathers both the annual flow of migrants and the decadal stock of migrants based on the data provided on "duration of residence." The migrant's duration of stay at place of residence is provided by the Census in four categories viz. (1) less than 1 year, (2) 1-4 years, (3) 5-9 years, (4) 10 years and above. The total stock of migrants in a census year is computed by summing up all these four categories (see Table-1). Intercensal stock of migrants is computed by summing up the first three categories of duration i.e. less than 1 year, 1-4 years and 5-9 years. Annual flow of migrants is computed by decomposition of the stocks in respective categories. The methodology of decomposition is as follows:
No. of migrants (M) in census year (y) :

$M(y)$ = migrants with duration < 1 year. ……….. (i)

No. of migrants (M) in year (y-t), $_{t=1,2,3,4}$:

$M(y\text{-}t)$ = (migrants with duration 1-4 years)/4 ……….. (ii)

No. of migrants (M) in year (y-t), $_{t=5,6,7,8,9}$:

$M(y\text{-}t)$ = (migrants with duration 5-9 years)/5 …… … (iii)

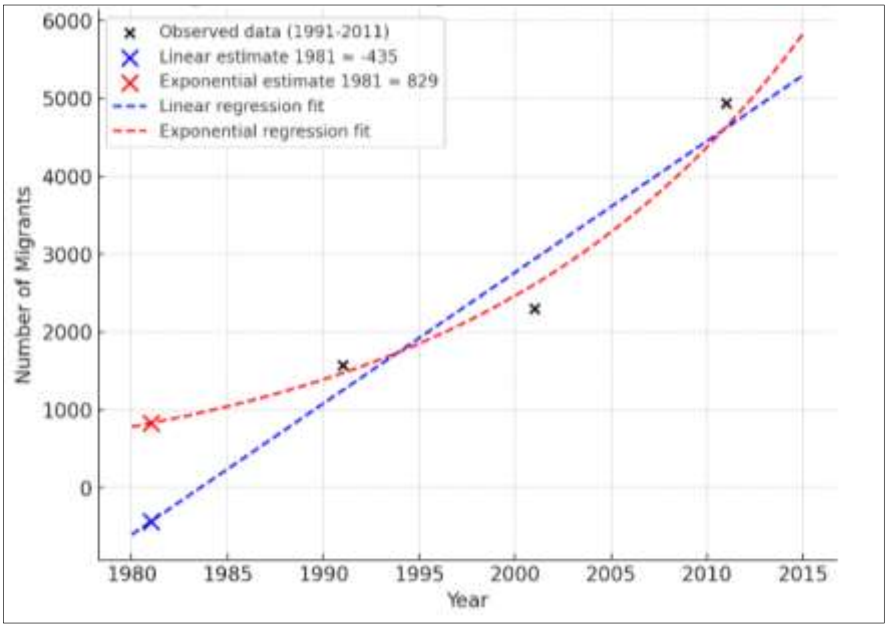
The annual migrant flows for the years 1982–2011 have been calculated retrospectively in this way (see Table 2). By fitting the annual data from 1991, 2001, and 2011 into an exponential regression fit, the result for 1981 is projected backwards (see result 2). There are two limitations to the data sources used in this study: the Census 1981 migration data is less trustworthy because Assam state did not conduct a census that year, and the yearly decomposed migration flows cannot precisely follow a smooth natural curve (see Figure 3).

Table-2: Decomposition of annual educational out-migration flows from Northeast states of India (1981-2011)

Year	Migrants	Year	Migrants	Year	Migrants
1981	829	1991	1569	2001	2296
1982	801	1992	839	2002	1056
1983	801	1993	839	2003	1056
1984	801	1994	839	2004	1056
1985	801	1995	839	2005	1056
1986	801	1996	839	2006	1056
1987	3078	1997	5042	2007	5630
1988	3078	1998	5042	2008	5630
1989	3078	1999	5042	2009	5630
1990	3078	2000	5042	2010	5630
				2011	4938

Source: Computed from Census of India 1991, 2001 & 2011, D-3 tables.

Figure-2: Estimation of annual migration in 1981

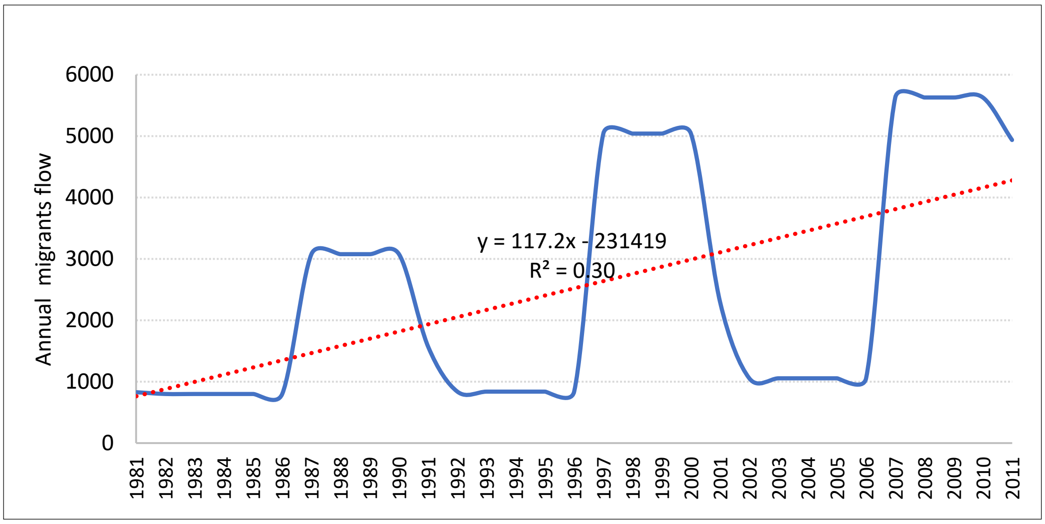


Source: Computed from Census of India 1991, 2001 & 2011.

Model fitting

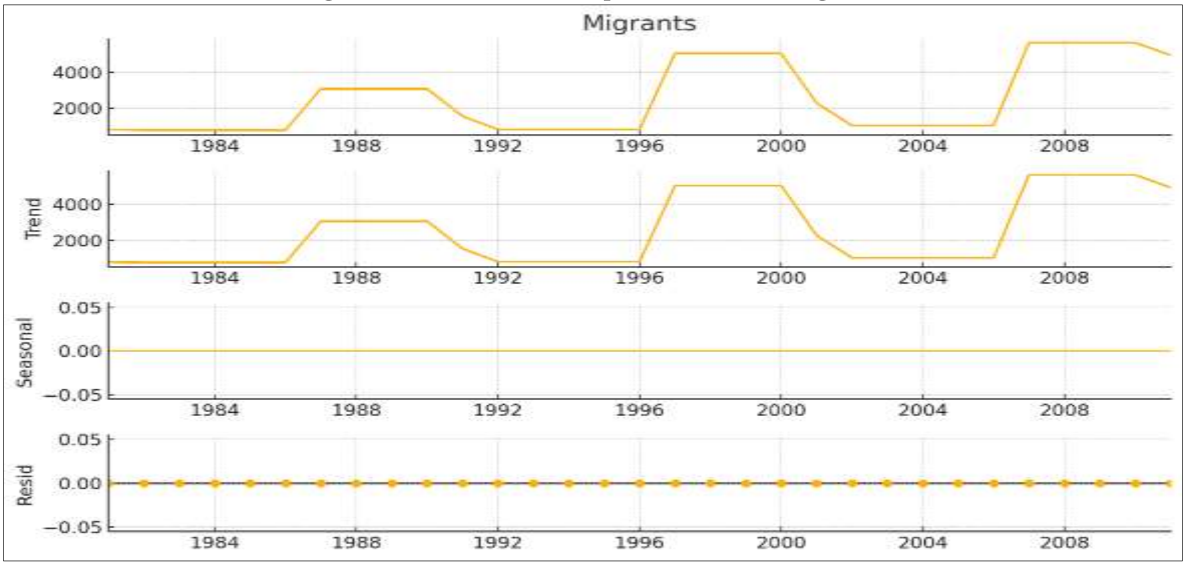
Because of the structure of the data, the annual educational out-migration flow data does not follow a smooth trendline (Figure-3). It is necessary to determine whether or not this data series contains seasonality before applying the ARIMA model. SARIMA models can better match data with seasonal behavior, but ARIMA models are appropriate in other cases. The *seasonal_decompose* function from the Python *statsmodels* module is used to check for seasonality. The seasonal decomposition analysis has revealed no seasonality in the data series. There is no recurring seasonal pattern throughout the measured years since the standard deviation of the seasonal component is zero. Only long-term trend changes are reflected since the time between observations is too long to catch seasonal or intra-annual variability. The dataset represents a non-seasonal time series showing long-term migration trends rather than seasonal variation (see figure-4).

Figure-3: Annual educational out-migrants flow from Northeast India



Source: Computed from Census of India 1991, 2001 & 2011.

Figure-4: Seasonal decomposition of out-migrants



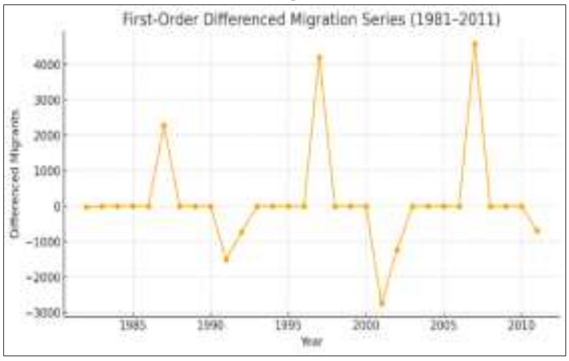
Another requirement for using the ARIMA model is that the data series must be stationary. For this reason, the Augmented Dickeddy-Fuller (ADF) test is used to determine whether or not the data series has stationarity. The non-stationarity of the data series is the ADF test's null hypothesis. This test uses the *adfuller* function from the *statsmodels* module in Python. Table 3 illustrates that the p-value (0.9229) is higher than the significance level (0.05) and the ADF test statistic (-0.3179) is greater than all crucial values. As a result, the null hypothesis cannot be rejected, and the educational out-migration data series is verified to be non-stationary.

First-order differencing is performed on the data series to make it stationary (see figure-5). The Python function of differencing is *diff( ).dropna( )*. After differencing, ADF test is again conducted on the differenced data series. Now, post-differencing test statistic (-7.5006) and p-value (0.000) confirm that the data series has been stationary and suitable for ARIMA modelling.

Table-3

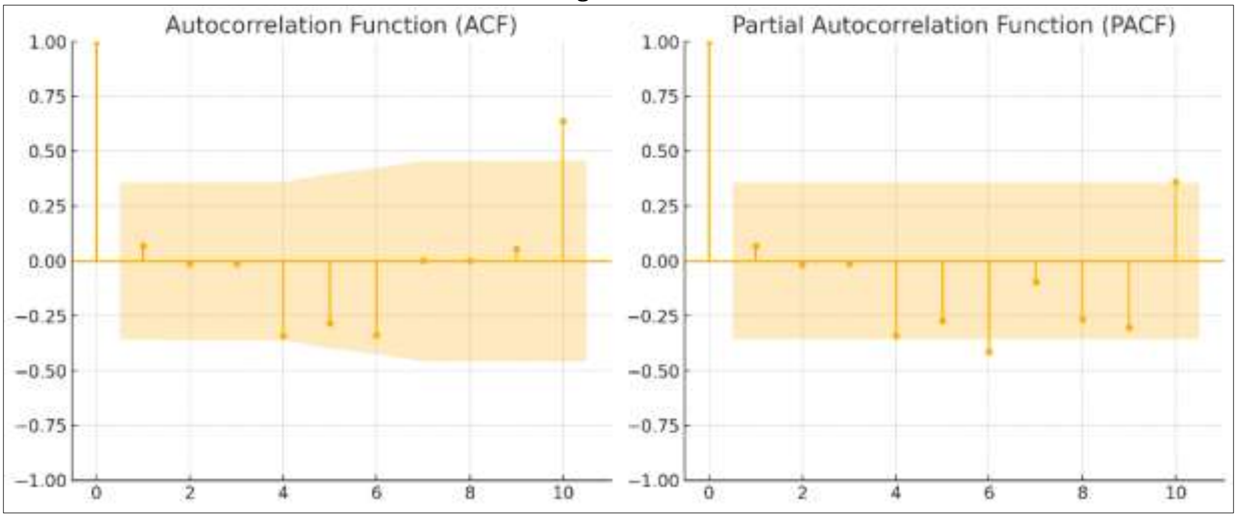
Augmented Dickey-Fuller (ADF) test results		
Metric	Value (pre-differencing)	Value (post-differencing)
ADF Statistic	–0.3179	-7.5006
p-value	0.9229	0.0000
Lags used	9	8
Observations	21	21
Critical value (1%)	-3.788	-3.788
Critical value (5%)	-3.013	-3.013
Critical value (10%)	-2.646	-2.646

Figure-5



Now, for fitting an ARIMA(p,d,q) model, three parameters i.e. *p*, *d*, *q* are required. 'd' is given by the differencing order i.e. d=1. 'p' and 'q' values are determined with the help of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). In Python, the ACF and PACF are plotted by using *plot_acf* and *plot_pacf* functions from *statsmodels* library (see figure-6). Based on the number of lag with significant spikes in the plots, a model of choice is **ARIMA(1,1,1)**
where, *p* = 1 (from PACF plot), *d* = 1 (from differencing order) and *q* = 1 (from ACF plot).
Once the model fit is ready, it can be put into machine learning to test forecast and that is how the model fit can be further improved in the next section of the article.

Figure-6



Machine Learning

In order for machine learning to identify trends and patterns and accurately predict future trends, the data series must be divided into two portions for training and testing. The ARIMA(1,1,1) model for train-test forecasting is now fitted with the previously differenced data series of the annual educational out-migrants flow (1981-2011). Test data spans the years 2002–2011, while training data spans the years 1981–2001. Figure 7(a) illustrates how the test forecast is unable to account for the variations in the data series. This model's negative R2 score suggests that the data series may be underfit or contain noise.

To find a better model fit, the ARIMA parameters i.e. *p, d, q* are optimized based on a full AIC grid search. This is done by employing the Auto-ARIMA function, which automates the process of finding the best fitting model parameters based on a statistical metric like the Akaike Information Criterion (AIC). In Python, Auto-ARIMA is executed by defining:

```
p = d = q = range(0, 4)
pdq_combinations = list(itertools.product(p, d, q))
```

and by AIC grid search for best (p,d,q):

```
model = ARIMA(y, order=param)
model_fit = model.fit()
results.append((param, model_fit.aic))
best_model_order = aic_results-iloc[0][')p,d.q)']
best_aic = aic_result.iloc[0]['AIC']
best_model = ARIMA(y, order=best_model_order)
best_fit = best_model.fit( )                              ……………………… (iv)
```

That is how ARIMA(0,2,1) is found with improved fit. The structure of the model is explained as follows:

p = 0, meaning no autoregressive (AR) component;

d = 2, meaning second-order differencing executed to achieve stationarity;

q = 1, meaning one moving-average (MA) term capturing short-term noise.

All things considered, this model soothes out annual swings while capturing a smooth long-term trend in migration. This model predicted a consistent upward trend with smaller errors (see table 4) (see image 7b). 15% of the test data's variance can be explained by it. To increase the model fitness, the entire machine learning and model fitting processes must be repeated as needed. In the following section of the article, the best-fitting models that have been discovered are created and described.

Figure-7(a)

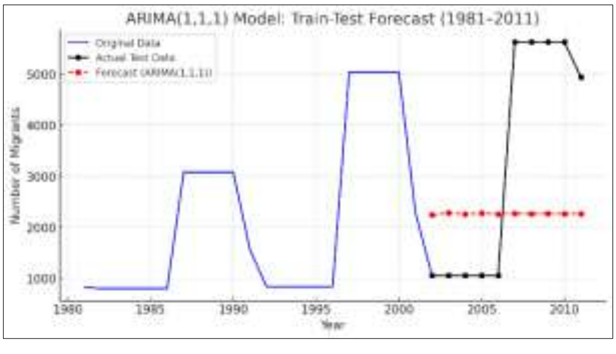


Figure-7(b)

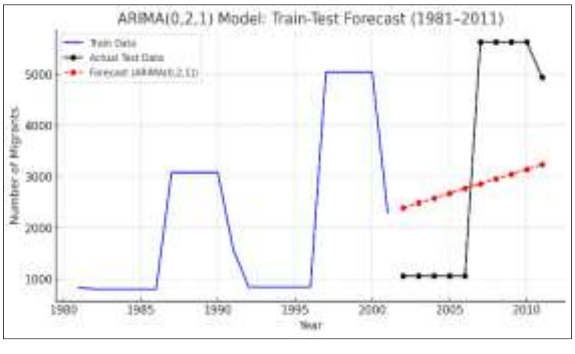


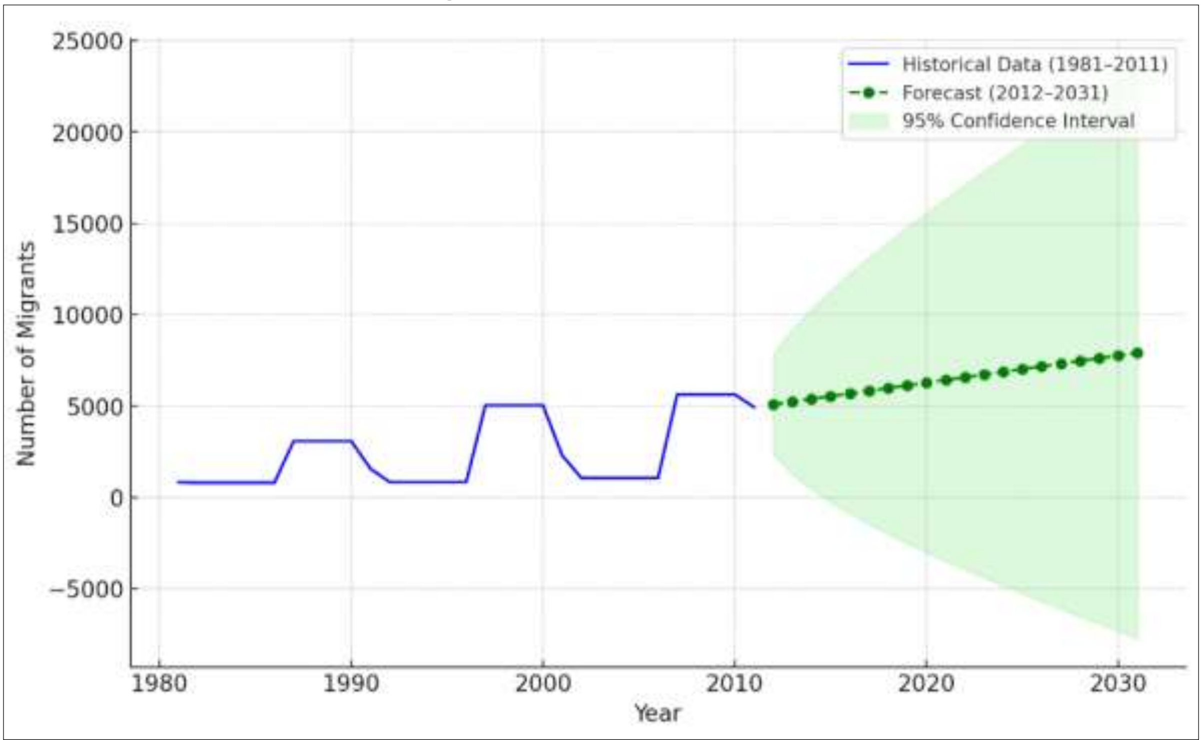
Table-4: ML-based ARIMA models training and testing results

Metrics	ARIMA(1,1,1)	ARIMA(0,2,1)
MAE	2215.98	1983.31
RMSE	2442.45	2056.40
R^2	-0.204	0.15
Source: Computed from Census of India 1991, 2001 & 2011.		

Findings of the forecast

Figure-8 displays the 95% confidence interval (CI) for the annual forecast of the flow of student out-migrants from the Northeastern states of India for the years 2012–2031. From about 5,000 migrants in 2012 to almost 8,000 migrants by 2031, the forecast indicates a consistent upward trajectory. The intercensal stock of migrants is expected to be 57,557 in 2021 and 72,423 in 2031 based on the sum of the annual flows (see table 5). As projection horizons increase, the widening CI shows growing uncertainty over time, which is common for ARIMA models. The pattern points to a sustained increase in educational outmigration, which is in line with long-term social and economic trends that encourage students to relocate outside of the area in order to pursue higher education.

Figure-8: Annual forecast (2012-2031) of student out-migrants from NE-India states using auto-ARIMA(0,2,1) model



The model's in-sample forecast typically deviates from actual values by roughly 750 migrants indicating a modest forecast accuracy, according to the mean absolute error (MAE) value. The root mean square error (RMSE) shows no extreme errors but sporadic higher variances. The model has a moderate goodness-of-fit, which is appropriate for long-term social data, and explains roughly 48% of the variance in the migration trend based on the R-squared score. Although the structure is usually stable, residuals show some autocorrelation, indicating that the model may not fully represent all time dependencies (Ljung-Box test p-value < 0.05). Although there are still short-term swings, the model fits long-term trend predicting rather well overall.

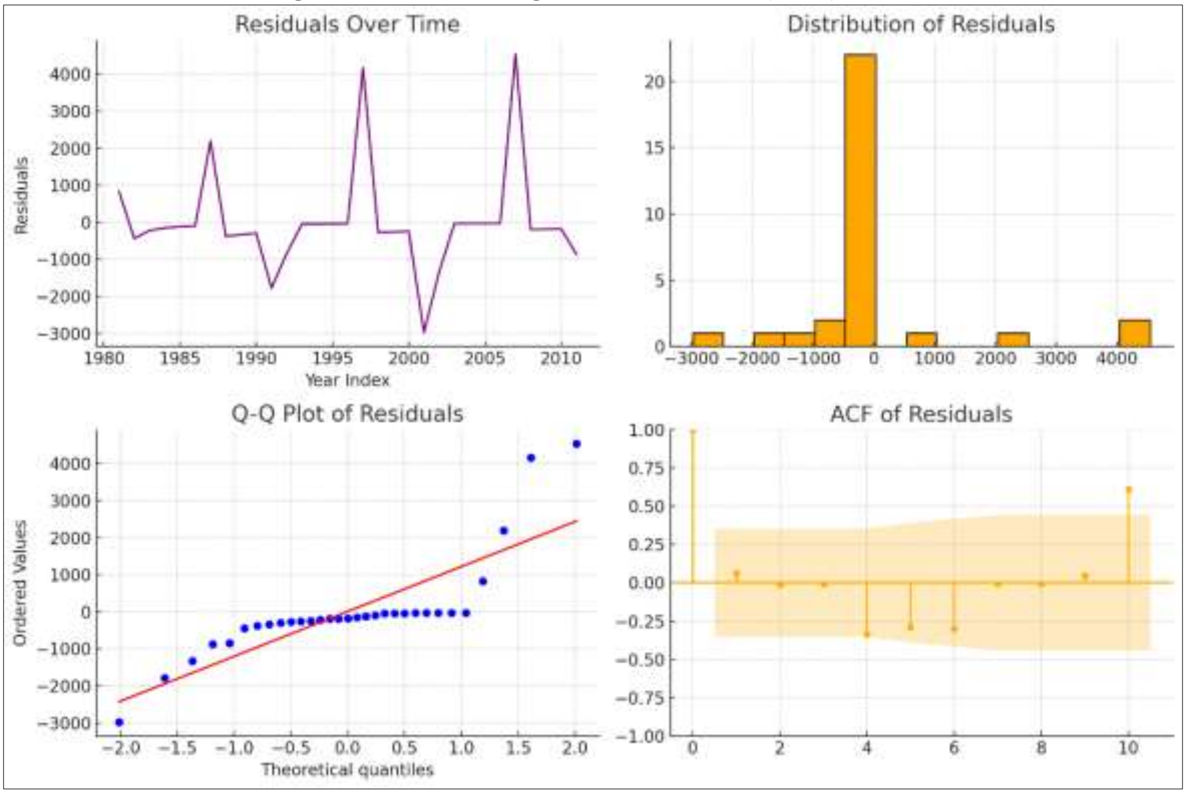
Table-5: Auto-ARIMA(0,2,1) forecast of annual student out-migrants from Northeast India states (2012-2031)

Forecast year	Migrants flow forecast	Intercensal stock forecast
2012	5087	57557
2013	5235	
2014	5384	
2015	5533	
2016	5681	
2017	5830	
2018	5979	
2019	6127	
2020	6276	
2021	6425	
2022	6573	72423
2023	6722	
2024	6871	
2025	7019	
2026	7168	
2027	7317	
2028	7465	
2029	7614	
2030	7763	
2031	7911	
Model fitness metrics		value
MAE		749.79
RMSE		1388.89
R^2		0.48
Ljung-Box test p-value		0.001

Source: Computed from Census of India 1991, 2001 & 2011.

The model's residuals are primarily random with brief spikes in error volatility. The model eliminates the primary trend as evidenced by the oscillations around zero. A few surges (between 1997 and 2000 and between 2006 and 2009) point to transient shocks or aberrant migration occurrences. The residuals histogram is somewhat skewed and generally bell-shaped suggesting modest non-normality. The majority of residuals cluster close to zero indicating minimal systematic bias. With sporadic significant deviations, errors are reasonably cantered which is appropriate for demographic and economic data. The majority of the points in the Q-Q plot of residuals follow the red 45° line but the tails slightly deviate, suggesting light-tailed non-normality. It implies that significant changes in migration (spikes and dips) were not adequately recorded. residuals that are somewhat non-normal but within reasonable bounds for ARIMA trend forecasting. With the exception of small lags, the majority of spikes lie inside the 95% confidence interval as shown by the ACF (Autocorrelation Function) of Residuals. It verifies that residuals are essentially uncorrelated, which is a prerequisite for an ARIMA fit to be considered legitimate. The residuals exhibit white noise behaviour suggesting that the systematic structure in the data is well-modelled by ARIMA(0,2,1).

Figure-9: Residual diagnostics of ARIMA(0,2,1) model



Conclusions & Recommendations

A consistent increase trend in student mobility between 2012 and 2031 is revealed by using the auto-ARIMA(0,2,1) model to forecast educational out-migration from Northeast India. According to the model, the annual student outflow will rise from about 5,000 in 2012 to over 8,000 by 2031 with intercensal stock estimates of 57,557 in 2021 and 72,423 in 2031. These findings highlight the persistence of educational migration as a structural phenomenon influenced by disparities in access to higher education opportunities and regional economic inequalities (McDuie-Ra, 2013; Nongkynrih, 2020). The model fitness metrics i.e., MAE (≈750), RMSE (≈1389) and R^2 (0.48) indicate a moderate but acceptable level of predictive accuracy for socio-demographic data which are typically subject to irregular fluctuations (Box & Jenkins, 1970; Hyndman & Athanasopoulos, 2018). The residual diagnostics demonstrate that the ARIMA(0,2,1) model successfully reflects the systematic temporal structure, despite the Ljung-Box test findings suggesting some residual autocorrelation. Consequently, the findings support the model's validity as a trustworthy instrument for medium- to long-term forecasting in migration research, especially in areas with limited data and patterns influenced by slow societal change (Zhang, 2003).

There are various policy ramifications to the anticipated rise in educational out-migration from Northeast India. First, it emphasizes how critical it is to increase the capacity of higher education in the Northeastern states by enhancing academic program diversity, teacher development, and institutional quality (Das & Saikia, 2019). The structural push forces that force students to leave the area could be lessened by bolstering local colleges and creating specialized centers of excellence. In order to retain or reintegrate skilled graduates, policy interventions such as targeted return-migration programs, scholarships, and regional employment incentives are required because prolonged out-migration implies a continuous loss of young human capital (McDuie-Ra, 2013; Nongkynrih, 2020). Lastly, policymakers may be able to predict changes in the demand for higher education and local labour markets by using machine learning and predictive analytics into educational planning to enable dynamic monitoring of migration trends (Makridakis et al., 2018; Murphy, 2012). State and national agencies can create flexible migration and education policies that support balanced regional development and provide fair access to higher education throughout India by utilizing evidence-based forecasting.

References

1. Amini, A., Gerostathopoulos, I., & Prehofer, C. (2018). Big data analytics architecture for real-time traffic control. *Journal of Big Data*, 5(1), 1–19.
2. Avinash, G., et al. (2025). Time series forecasting of bed occupancy in mental health facilities in India using machine learning. *Scientific Reports*, 15(2686), 1–12. https://doi.org/10.1038/s41598-025-86418-9
3. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer.
4. Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.
5. Chyrmang, R. (2011). Magnitude of Migration from North-Eastern Region of India. In S. Irudaya Rajan (Ed.), *Migration, Identity and Conflict – India Migration Report 2011*. New Delhi: Routledge.
6. Das, K., & Saikia, A. (2019). Educational migration and regional development in Northeast India. *Asian Journal of Migration,* 6(1), 83–101.
7. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning.* Springer.
8. Hussain, N. H. M., et al. (2021). Machine learning of the reverse migration models for population prediction: A review. *Turkish Journal of Computer and Mathematics Education*, 12(5), 1830–1838.
9. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). Melbourne, Australia: Texts.
10. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.
11. Lusome, R., & Bhagat, R. B. (2020). Migration in Northeast India: Inflows, outflows and reverse flows during pandemic. *The Indian Journal of Labour Economics*, 63, 1125-1141.
12. Lyndem, B. and De, U. K. (Eds.) (2004). *Education in North-East India: Experience and Challenge.* New Delhi: Concept Publishing Company.
13. Makridakis, S., Spilliotis, E. & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting,* 34(4):802-808.
14. McDuie-Ra, D. (2012). The ‘NorthEast’ Map of Delhi. *Economic and Political Weekly*, 47(30):69-77.
15. McDuie-Ra, D. (2013). *Northeast Migrants in Delhi: Race, Refuge and Retail.* Amsterdam University Press.
16. Mistri, A and Sardar, S. S. (2022). Student Migration from the North-East India: Level, Trend, Pattern and Challenges. *Demography India*, 51(1):40-62.
17. Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective.* MIT Press.
18. Nongkynrih, A.K. (2020). Migration for higher education from Northeast India: Trends and implications. *Economic and Political Weekly,* 55(36), 45–52.

19. Padmaja, P., et al. (2024). Prediction of employee attrition using machine learning algorithm. *Journal of Computational Analysis and Applications*, 33(5), 616–620.
20. Shimray, U. A., & Usha Devi, M. D. (2009). Trends and patterns of migration: Interface with education – A case of the North-Eastern Region. (*Social and Economic Change Monographs*, No. 15). Bangalore: Institute for Social and Economic Change (ISEC).
21. Vatti, P.R., Vatti, V.R. & Sen, K. (2024). Advancing Data Science with System Intelligence: A Machine Learning Approach to Predictive Data Engineering. *Journal of Computational Analysis and Applications (JoCAAA)*, *33*(08), 2344–2360. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/2086
22. Wei, W.W.S. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. Pearson.
23. Yang, S., et al. (2020). Forecasting outbound student mobility: A machine learning approach. *PLOS ONE*, 15(9), e0238129. https://doi.org/10.1371/journal.pone.0238129
24. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0