International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

New-Age Framework for Educational Reforms

Mehta Twisha Piyush

Student, GIIS

Abstract

This research describes a framework for utilizing machine learning and data visualization with publicly available education data to create educational reforms. This study aims to address the systemic issues hindering equal learning opportunities between developed and developing countries by analyzing key global education indicators. With the help of clustering and predictive analysis, the work reveals significant factors affecting education, including but not limited to the completion rates, tertiary education enrollment, and literacy rates. The framework also leverages social media and machine learning to drive educational reforms and reduce policy implementation lag by informing policymakers.

Keywords: Education, Machine Learning, Reform, Framework, Predictive Analytics, Data-Driven Education Reforms, Policy Implementation

1. Introduction

In today's rapidly evolving world, the gap between developed and developing nations continues to widen, creating stark inequalities across multiple sectors, including education. Education, widely regarded as a cornerstone of economic development and social progress, is unequally distributed, leaving millions in developing nations with inadequate access to quality learning. This inequality perpetuates a vicious cycle of poverty, limited opportunities, and underdevelopment. While traditional frameworks for educational reform have made strides in addressing some of these systemic issues, they have often fallen short due to **slow policy implementation, a lack of real-time data, and an inability to scale solutions effectively.** For instance, a UNESCO report highlighted that education systems in low-income countries experience an average policy implementation lag of 5-10 years, which severely hampers the ability to respond to urgent needs and adapt to rapidly changing environments.

The advent of machine learning (ML) and artificial intelligence (AI) offers unprecedented opportunities to reshape how we approach educational reform. By leveraging large-scale educational datasets and real-time insights, these technologies can identify patterns, predict outcomes, and recommend targeted interventions. Moreover, social media has emerged as a powerful tool for advocacy, enabling faster dissemination of knowledge and mobilizing communities to influence policymakers, as seen in campaigns such as Malala Fund's #YesAllGirls initiative, which rapidly gathered global attention to push for girls' education reforms.

This paper proposes a new-age framework that harnesses the power of ML, AI, and social media to drive educational reforms, offering a data-driven, scalable, and effective solution to the systemic challenges. By addressing issues such as completion rates, literacy, and educational enrollment through dynamic clustering and predictive analysis, this framework not only accelerates reform efforts but also enables policymakers to identify which countries require the most support and how their interventions should be tailored. The framework provides insights into country-specific educational needs, allowing for

customized policies that adapt to each nation's unique challenges and current situation, ensuring that reforms are both impactful and sustainable in the long term.

2. Literature Review

In today's world, stark inequalities persist between developed and developing nations. Developed nations enjoy a high quality of life, while developing nations struggle with numerous challenges. Among these, income inequality remains a significant concern, and its effects ripple across various sectors, especially education. According to the World Inequality Report, global inequality has increased since 1980, with the top 1% capturing twice as much global income growth as the bottom 50%. For example, the average income of individuals living in North America is 16 times higher than that of people in sub-Saharan Africa. This disparity adversely impacts access to quality education, as well-funded educational systems are often a luxury in developing regions. The resulting inequality in education perpetuates a vicious cycle, where those without access to quality education are less likely to break free from poverty (World Inequality Report, 2020).

Education, widely recognized as a critical tool for empowerment, provides individuals with opportunities to improve their circumstances and contribute to societal development. Economists Richard Murnane and Greg Duncan found that between 1972 and 2006, high-income families in the United States increased their spending on their children's enrichment activities by 150%, while low-income families increased their spending by only 57%. This stark difference reflects the understanding that educational success is crucial in today's knowledge-driven economy (Murnane & Duncan, 2011).

Sociologist Marianne Cooper, in her book *Cut Adrift*, further emphasizes the increasing importance of education in achieving economic stability (Cooper, 2014). However, unequal educational resources, especially in developing countries, prevent education from becoming a mainstream tool for empowerment. Quality differences between schools have a dramatic impact on productivity and national growth rates, as demonstrated by Hanushek and Kimko (2000), further exacerbating the cycle of poverty in these regions. Addressing educational inequality is therefore essential for uplifting the global quality of life. Research by Abdul Abdullah and Hristos Doucouliagos (2013) has shown that education not only reduces the income share of top earners but also increases the share of income among the bottom earners, indicating the redistributive power of education. However, systemic challenges such as inadequate infrastructure, insufficient funding, and political instability hinder the development of effective educational systems in many developing nations. These barriers must be addressed through significant policy changes to empower people via education.

Policy makers hold the key to addressing these systemic educational challenges. Policies focused on improving characteristics such as class size, educational facilities, teacher qualifications, and community investment levels can substantially reform educational systems. However, traditional reform efforts are often hindered by slow bureaucratic processes and local advocacy efforts. This results in an implementation lag—where the delay between recognizing a problem and implementing a solution may diminish the effectiveness of the policy or even lead to counterproductive outcomes (Hanushek & Kimko, 2000). Accelerating policy changes is crucial, and compelling policy makers to act with urgency can mitigate the lag between identifying problems and addressing them.

In recent years, emerging technological trends have opened new avenues for social impact. Machine learning (ML) and artificial intelligence (AI) now provide powerful tools for analyzing educational data and generating actionable insights. Moreover, governance changes have made large datasets from

developing nations publicly available, creating opportunities for detailed analysis and targeted interventions. Social media, with its widespread reach and influence, has also surpassed traditional media as a platform for advocacy, offering new ways to engage both the public and policy makers.

Given these developments, the new-age framework proposed in this paper builds on traditional educational reform models by integrating modern technological advancements. Earlier frameworks, such as Human Capital Theory, emphasized education's role in improving economic outcomes but lacked the real-time data and insights necessary for timely policy adjustments. Traditional approaches, while focusing on infrastructure improvements and teacher quality, were limited by slow feedback loops between policy implementation and evaluation.

This new framework leverages machine learning and artificial intelligence to enable real-time analysis of educational data, allowing for faster and more accurate identification of challenges. The use of clustering algorithms and predictive analytics adds a dynamic dimension to educational reform, providing insights into which countries or regions require the most support and how policies can be tailored based on current educational indicators.

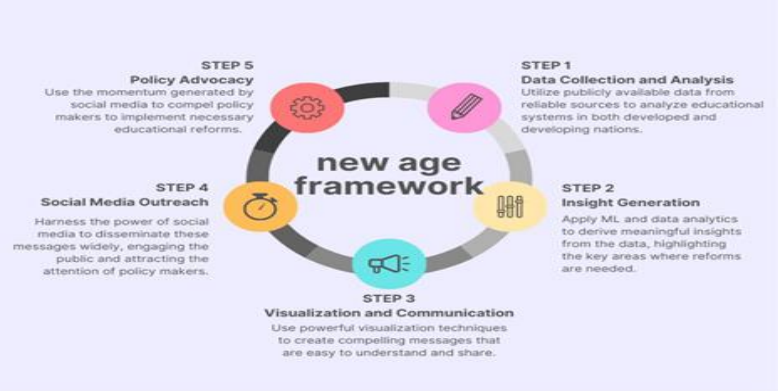
Additionally, integrating social media as an advocacy tool reduces policy implementation lags by raising public awareness and directly engaging policy makers in real-time conversations. This method offers a significant improvement over slower traditional advocacy models, which often failed to maintain momentum. By combining data-driven insights with widespread social mobilization, this framework marks a substantial advancement over earlier educational reform efforts that struggled with delays and insufficient actionable data (Hanushek & Kimko, 2000; Doucouliagos & Abdullah, 2013).

3. New-Age Framework

The framework involves the following steps:

1. Data Collection and Analysis**:** Utilize publicly available data from reliable sources to analyze educational systems in nations.
2. Insight Generation**:** Apply machine learning and data analytics to derive meaningful insights from the data, highlighting the key areas where reforms are needed.
3. Visualization and Communication**:** Use powerful visualization techniques to create compelling messages that are easy to understand and share.
4. Social Media Outreach**:** Harness the power of social media to disseminate these messages widely, engaging the public and attracting the attention of policy makers.
5. Policy Advocacy: Use the momentum generated by social media to compel policy makers to implement necessary educational reforms.

Figure 1: Visual Representation of Framework



4. Implementation of Framework

4.1 Data Collection and Analysis

Dataset is taken from a publicly available dataset https://www.kaggle.com/datasets/nelgiriyewithana/world-educational-data/data

The "World Education Data" dataset offers a comprehensive global view of education. It covers essential metrics like out-of-school rates, completion rates, proficiency levels, literacy rates, birth rates, and enrolment in primary and tertiary education. These metrics help us to draw valuable insights from the dataset. There is a total of 202 entries and 29 columns.

All data analytics done can be viewed at: https://github.com/TwishaMehta/NeurIPS_2024

Before analyzing the data, some data pre-processing needs to be performed on it to ensure its quality.

A check for missing values in the dataset revealed none initially, but upon further inspection, it was found that '0' acted as a placeholder for missing data in certain fields. To address these '0' values, I employed **regression imputation**—a method that leverages relationships between variables within the same country to predict missing values. This approach ensures that imputed values are consistent with the country's overall educational profile, which is crucial given the disparities between countries in the dataset.

For instance, missing literacy rates were predicted using a regression model based on enrollment rates, completion rates, and proficiency scores. **Regression imputation** was chosen over simpler methods such as mean or median imputation because those methods, which replace missing values with global averages, would fail to capture the unique educational profiles of different countries. Imputing based on these averages would distort the data, especially in cases where countries exhibit extreme or distinctive educational indicators.

Similarly, **K-Nearest Neighbors (KNN) imputation** was considered but ultimately deemed inappropriate. KNN relies on similarities between neighboring data points, but due to the vast differences between developed and developing nations in this dataset, relying on neighboring countries for imputation could introduce significant bias and misrepresent the unique challenges each country faces.

By contrast, **regression imputation** allowed for country-specific relationships between variables—such as literacy rates and enrollment patterns—to be captured more effectively. This approach preserves the integrity of each country's data and ensures that imputed values are contextually relevant, offering a more accurate representation of the dataset as a whole. In doing so, the method enhances the quality of the analysis, maintaining consistency with the country's demographic and educational characteristics.

After data pre-processing was done, a preliminary exploration of the dataset was conducted to assess its structure and completeness. The data.describe() function was used to compute key descriptive statistics for the dataset. These statistics reveal significant insights into global education trends.

Figure 2: Overview of Basic Statistics of Raw Data

	Latitude	Longitude	OOSR_Pre0Primary_Age_Male	OOSR_Pre0Primary_Age_Female	OOSR_Primary_Age_Male	OOSR_Primary_Age_Female	OOS
count	202.000000	202.000000	202.000000	202.000000	202.000000	202.000000	
mean	25.081422	55.166928	19.658416	19.282178	5.282178	5.569307	
std	16.813639	45.976287	25.007604	25.171147	9.396442	10.383092	
min	0.023559	0.824782	0.000000	0.000000	0.000000	0.000000	
25%	11.685062	18.665678	0.000000	0.000000	0.000000	0.000000	
50%	21.207861	43.518091	9.000000	7.000000	1.000000	1.000000	
75%	39.901792	77.684945	31.000000	30.000000	6.000000	6.750000	
max	64.963051	178.065032	96.000000	96.000000	58.000000	67.000000	

8 rows × 28 columns

For example:

- The **mean youth literacy rate** for males is 57.28%, whereas for females, it is significantly lower at 35.08%. This already highlights the gender disparity present in the field of education.
- The **gross primary education enrolment rate** has a mean of 94.94%, with some countries achieving over 100% enrolment due to late or early enrolments, while the **gross tertiary education enrolment rate** is much lower at an average of 34.39%. This shows that the access to higher education remains a challenge in many regions.
- Furthermore, the **standard deviations** of key variables such as gross enrollment rates and literacy rates also reveal the wide disparity in educational outcomes between countries. The standard deviation for male youth literacy is 35.80%, while for females, it is 45.25%, which suggests that certain regions are disproportionately affected by poor educational infrastructure and policies, making **them prime targets for focused interventions**.

These preliminary analyses offer valuable context for the subsequent application of machine learning models, such as clustering algorithms, to group countries based on their educational performance, and regression models to predict key outcomes such as literacy and enrolment rates.

Exploratory Data Analysis (EDA) was undertaken to uncover relationships and patterns in the dataset and guide further analysis. Several techniques were employed, including the **Apriori Algorithm**, **Principal Component Analysis (PCA)**, and **t-SNE** (t-distributed Stochastic Neighbor Embedding), each offering unique insights into the structure of the data.

To begin, isolating the numerical columns and generating a **correlation matrix** helped identify relationships between various educational and economic factors. The Pearson correlation coefficients were calculated for all numerical variables, revealing several important relationships:

- It showed a **positive** correlation between primary school completion rates and subsequent secondary school completion rates, suggesting that improvements in primary education have a ripple effect across other educational levels.
- Out-of-school rates (OOSR) exhibited a **negative** correlation with literacy rates, highlighting the adverse effects of students dropping out of school.

Next, association rule learning using the Apriori Algorithm discovered interesting associations between variables such as literacy rate, tertiary education enrolment, and unemployment rate. For example, the following key associations were identified:

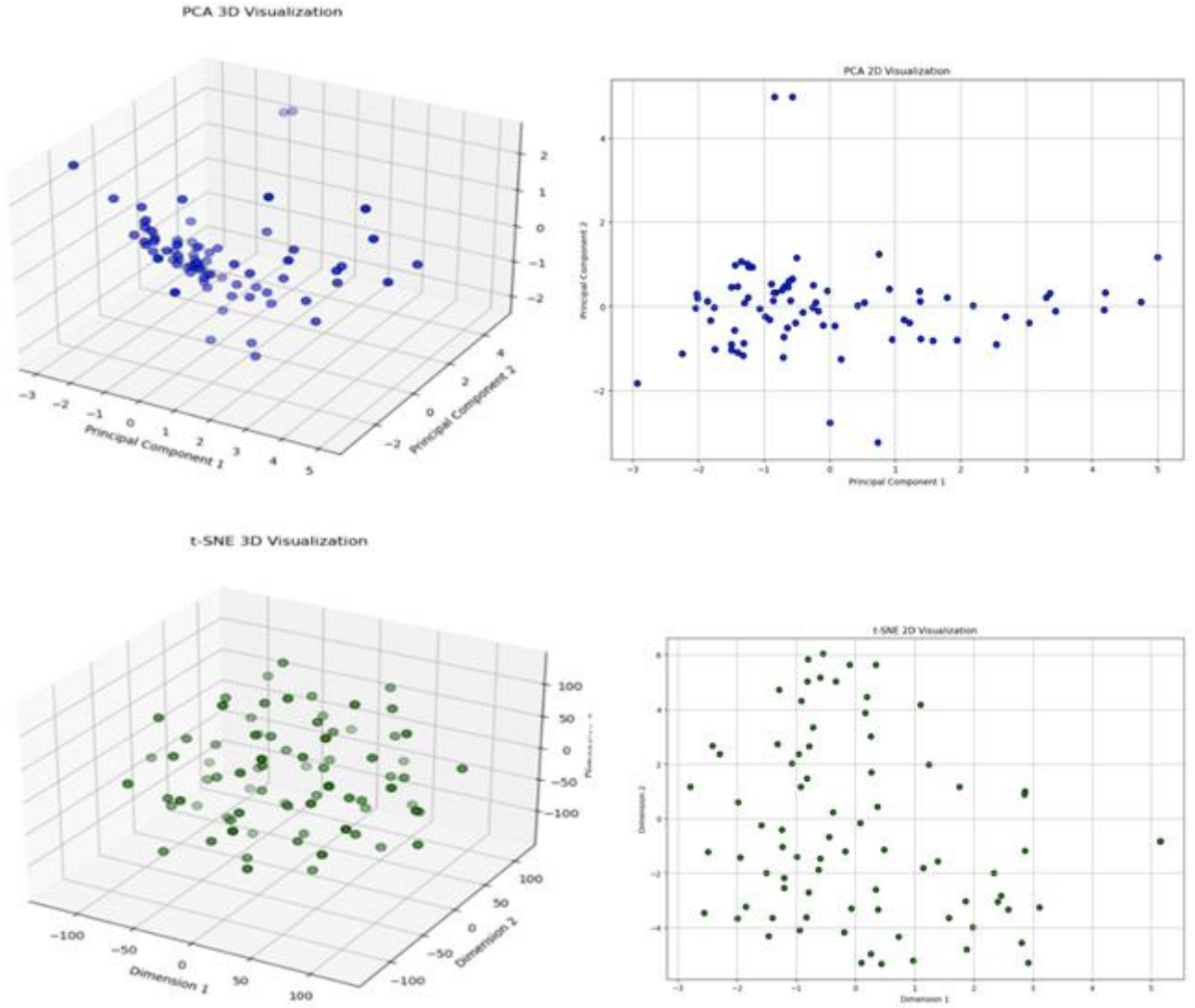
1. **High female youth literacy rate is associated with lower birth rates**, with a support of 31.65%, a confidence of 37.31%, and a lift of 1.18.
2. **Low unemployment rates are associated with low tertiary education enrollment**, with a support of 48.10%, a confidence of 71.70%, and a lift of 1.07.

The Apriori algorithm helped reveal these associations, offering insights into how educational and economic factors may co-occur, providing potential areas for policy intervention. For example, improving female literacy could help lower birth rates, contributing to better economic outcomes for developing nations.

Next, PCA analysis results obtained from the current set of variables highlight the fact that countries with similar values of education indices and dissimilar profiles are grouped together on the 2D plot. The 3D plot resolves the issue with identifying the clustering and offers the option to indicate the variances that give the three biggest groupings while also differentiating overlying clusters from the two-dimensional

plot. t-SNE depicts the non-linear relationships of the countries' educational metrics. This can be seen in Figure 1 and 2.

Figure 3: 2D and 3D Visualization of PCA and t-SNE



4.2 Insight Generation

After performing data preprocessing and data analysis, various machine learning models were trained to generate insights. These models, including **K-Means Clustering** and **Random Forest Regression**, were used to explore relationships between educational indicators and to predict factors influencing educational completion rates and enrolment rates.

4.2.1 K-Means Clustering

I used **K-Means Clustering** to explore patterns in educational indicators across countries, grouping them based on shared characteristics. Clustering was particularly useful because it allowed me to identify natural groupings of countries without needing predefined labels, unlike regression or classification models. This method helped reveal which countries face similar educational challenges, allowing for *tailored* policy interventions.

Before applying K-Means, I standardized the data to ensure that all features contributed equally, as K-Means is sensitive to differences in scale. This ensured that all educational indicators contributed equally to the clustering process. I applied the algorithm with 5 clusters, assigning each country to the cluster with the nearest mean. The resulting clusters were visualized using a scatter plot (Figure 4) of gross primary education enrolment versus gross tertiary education enrolment, with countries color-coded by cluster. I also saved the country-cluster associations in a CSV file for further analysis.

To dive deeper into the clustering results, I created a pair plot (Figure 5) to visualize the relationships between key educational indicators—such as literacy rates and enrolment rates—across the clusters. This plot showed clear trends: countries in lighter clusters (0 and 1) typically had lower enrolment and literacy rates, while those in darker clusters (3 and 4) had stronger educational systems with higher values across these indicators. A bar plot (Figure 6) showed the distribution of countries across clusters, with Clusters 1 and 2 containing the largest number of countries.

The analysis highlighted significant educational disparities. Clusters 0 and 1 represent countries with weaker educational systems, where reforms should focus on increasing access to basic education. In contrast, Clusters 3 and 4 represent countries with more developed systems, where the focus should be on enhancing quality and higher education accessibility. The positive correlation between primary and tertiary enrolment emphasizes the importance of early-stage educational investment, while the high correlation between male and female literacy rates suggests that improving literacy benefits both genders equally.

This clustering analysis provides valuable insights for designing policy interventions tailored to the educational needs of each group of countries. Countries in lower-performing clusters need foundational reforms, while those in higher-performing clusters can focus on quality enhancements and technological integration in education.

Figure 4: Graph showing the K-Means Clustering done

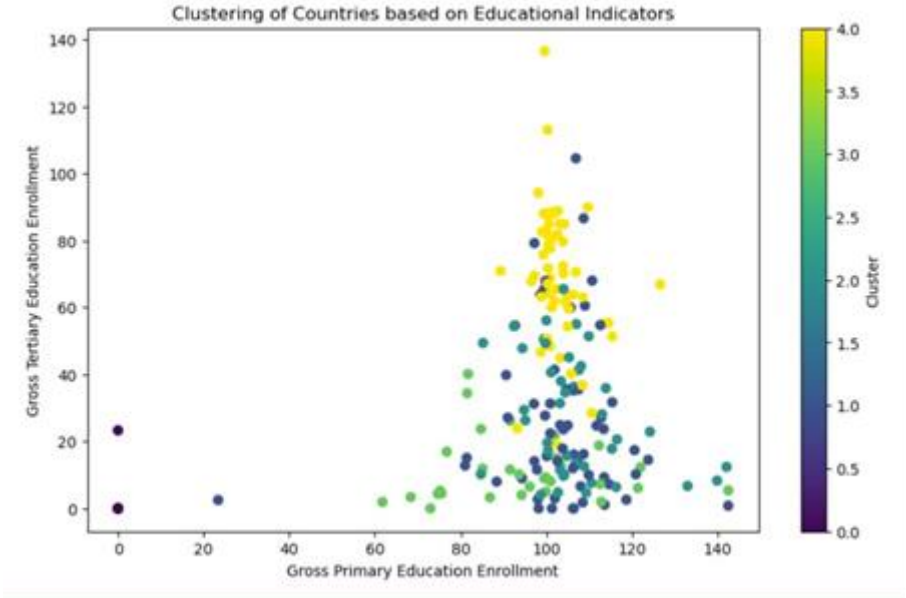


Figure 5: Pairplot of Educational Indicators by Clusters

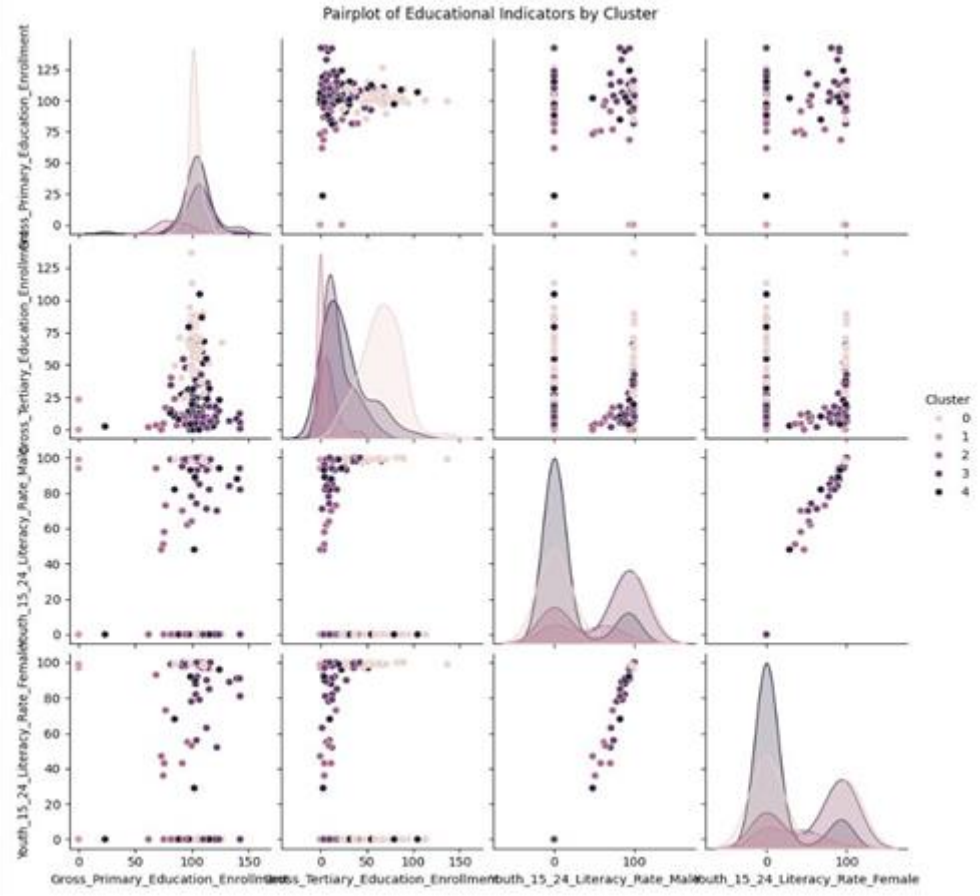
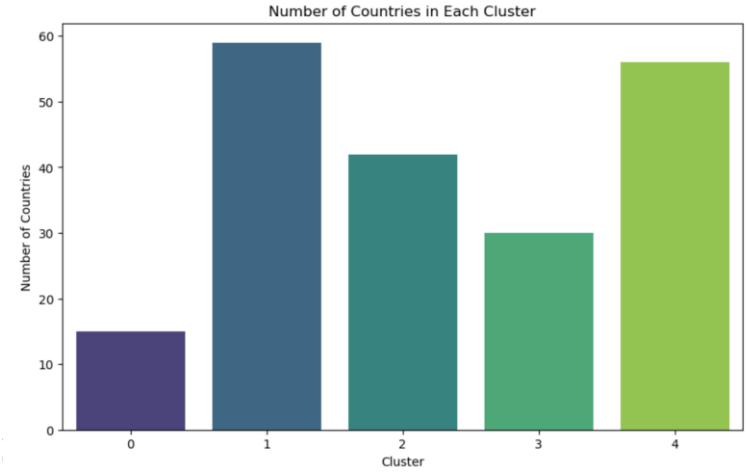


Figure 6: Bar Chart



4.2.2 Predictive Regression Model

I initially applied a predictive regression model to analyze the dataset. However, the model yielded poor results, with a high Mean Squared Error (MSE) and a low R-squared value. These metrics indicated that the model was ineffective at predicting literacy rates, with very limited explanatory power.

To address this, I tested three different models: Ridge Regression (linear regression with L2 regularization), Lasso Regression (linear regression with L1 regularization), and Random Forest

Regression (a non-linear model that captures complex relationships). Unfortunately, all three models produced similar results, leading me to realize that **the issue lay in my feature selection**.

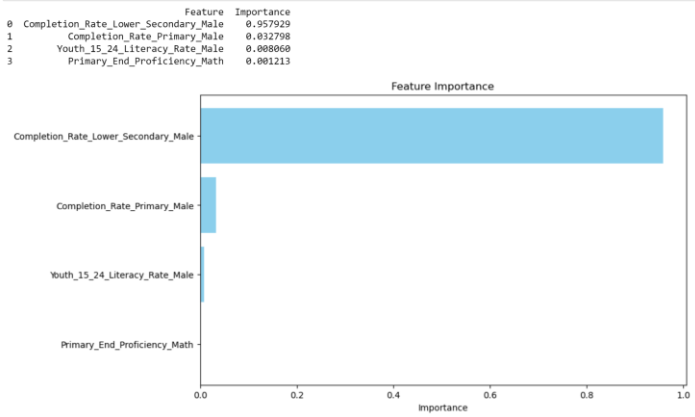
Feature selection is crucial in determining which variables are most predictive of the target outcome. To refine the selection process, I examined continuous variables that could serve as meaningful target variables based on the summary statistics. I focused on variables representing rates or percentages, such as Completion Rates (Primary, Lower Secondary, Upper Secondary), Proficiency Scores (Reading and Math), Youth Literacy Rates, Birth Rate, Gross Education Enrollments, and Unemployment Rate.

For the initial regression analysis, I selected **Completion Rate for Upper Secondary Males** as the target variable. A correlation matrix was computed to identify potential predictors. The analysis revealed strong positive correlations with completion rates at other educational levels, particularly Completion Rate for Lower Secondary Males and Females, suggesting that general educational attainment tends to progress similarly across genders and stages within countries. However, there were also moderate correlations with Unemployment Rate and Youth Literacy Rates, though these were weaker predictors.

I then set up a Random Forest Regression model to predict the completion rate for upper secondary males. The model performed significantly better, with an MSE of 147.37 and an R-squared value of 0.835, compared to the previous model's poor performance. To further enhance the model's performance, I used grid search to optimize hyperparameters, including the number of trees, maximum tree depth, and minimum samples required to split a node. The best model had an MSE of 72.99 (cross-validated) and a test MSE of 148.78, with an R-squared of 0.833.

Analyzing the feature importance scores from the Random Forest model revealed that Completion Rate for Lower Secondary Males was by far the most influential predictor, indicating that completion rates across educational levels are highly interrelated.

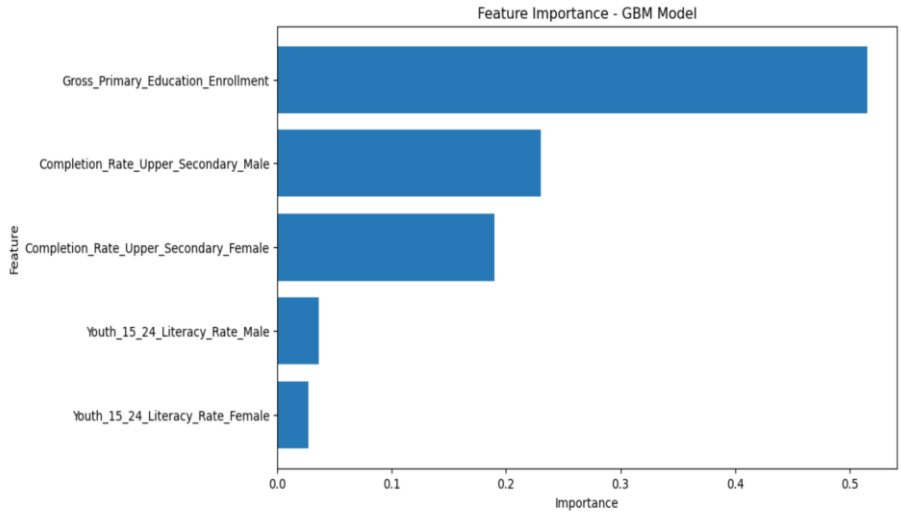
Figure 7: Feature Importance Graph for 'Completion Rate for Upper Secondary Males'



This suggests that interventions aimed at improving lower secondary education could have a significant impact on upper secondary completion rates. Other factors like primary completion rates and youth literacy rates, while less impactful, still play a role.

Next, I applied the same methodology with a different target variable: Gross Tertiary Education Enrollment. The feature importance analysis showed that Gross Primary Education Enrollment was the most significant predictor, underscoring the importance of early education as a foundation for higher educational attainment. Completion Rate for Upper Secondary Males and Females were also important predictors, reinforcing the importance of secondary education in the transition to tertiary education.

Figure 8: Feature Importance for 'Gross Tertiary Education Enrollment'



Interestingly, while Youth Literacy Rates were found to have a lower impact on tertiary enrollment, basic literacy remains a necessary foundation for academic success. The relatively lower influence of literacy rates may indicate that factors like completion rates and enrollment metrics are more immediate determinants of tertiary education participation.

I followed the same logic for female literacy rate and the top features were 'Uppersecondary-agefemale' with 36.24% and 'lower_secondary_age_female' with 33.23%. The out of school rates significantly impact the female literacy rates and policies should focus on reducing this and increasing enrolment in secondary education.

Some policy recommendations and insights generated from the above model:

1. **Strengthen Primary Education**: The strong relationship between primary enrollment and tertiary education highlights the need for investments in primary education. Improving the quality and accessibility of primary education—through better teacher training, reduced student-to-teacher ratios, and technology integration—will likely yield long-term benefits for tertiary enrollment.
2. **Support for Secondary Education Completion**: The high importance of completion rates at the secondary level suggests that targeted interventions to reduce dropout rates are crucial. Policies that provide academic support, career counseling, and socio-emotional learning can help students transition from lower to upper secondary education, and eventually to tertiary education.
3. **Gender-Specific Educational Policies**: Both male and female secondary completion rates were strong predictors of tertiary enrollment, indicating the need for gender-specific policies to ensure equal educational opportunities and support for both genders at critical stages.
4. **Literacy Programs**: Although literacy rates were less significant in predicting tertiary enrollment directly, they remain essential for academic success. Policies should continue to focus on improving literacy rates as part of a holistic approach to educational reform.
5. **Data-Driven Education Reforms**: This analysis highlights the value of using data analytics to inform educational policy. By identifying the most significant predictors of educational outcomes, policymakers can allocate resources more effectively and design interventions that address the specific needs of different educational levels.

The insights from this analysis underscore the critical role of early and secondary education in shaping tertiary enrollment. By focusing on these foundational stages, policymakers can drive improvements in

educational outcomes, thereby enhancing the socioeconomic prospects of the population. Data-driven reforms that address key predictors—such as primary and secondary completion rates—will be crucial in creating a more equitable and effective global education system.

4.3 Visualization and Communication

I created a blog post publishing the same results which can be read here – "twisha.net". However, the initial blog post did not generate significant traction, leading to the decision to pivot to a more visually engaging infographic strategy.

Hence, I created an infographic publishing the above results in an attractive way to enhance the outreach and understanding of these insights. The infographic visually represents the following:

- Key Data Insights: Visuals such as bar charts and pie charts depicting literacy rates, enrolment rates, and their impacts on socioeconomic factors.
- Clustering Results: A visual representation of the clusters, showing countries grouped by educational performance and the specific areas where reforms are needed.
- Predictive Models: Graphs showing the feature importance for predicting educational completion and enrolment rates, making it clear which factors are most influential.

4.4 Social Media Outreach

Based on my research, I identified that LinkedIn is a good platform for activism and Instagram is a good platform for reaching young audiences. Social media provides opportunity for community and gaining momentum for a cause. Publishing my infographic on these platforms garnered views and interactivity.

4.5 Policy Advocacy

The combined approach of data analysis, visualization, and social media outreach is designed to capture the attention of educational policymakers. The message created can reach professionals who in turn can buy in the idea and become an advocate to influence policymakers. By presenting clear, data-driven insights and engaging content, I aim to influence policy decisions, reduce implementation lag, raise public awareness and drive systemic changes.

5. Conclusion

This research framework harnesses the power of machine learning and data visualization to analyze key educational factors such as literacy and enrollment rates, providing data-driven insights to guide educational reforms. By generating targeted policy recommendations, the framework addresses specific country needs, while also reducing policy implementation delays. Through enhanced outreach via social media, it effectively mobilizes communities and influences policymakers more rapidly. The findings emphasize the critical role of foundational education in improving overall educational outcomes and addressing global inequalities. Looking ahead, this framework has the potential to incorporate real-time policy feedback loops and further integrate AI-driven tools to make educational reforms even more dynamic, responsive, and impactful on a global scale.

6. References

1. Abdhullah, A., Doucouliagos, H. & Manning, E., 2013. Does Education Reduce Income Inequality? A Meta-Regression Analysis. *Journal of Economic Surveys,* 29(2), pp. 301-316.

2. Bradley, S. & Green, C., 2020. *The Economics of Education.* 2 ed. s.l.:s.n.
3. Cooper, M., 2014. *Cut adrift: Families in insecure times,* s.l.: University of California pres.
4. Duncan, D. & Murnane, R., 2011. *Whither opportunity?: Rising inequality, schools and children's life chances,* s.l.: Russell Sage Foundation.
5. Hanushek, E. A. & Kimko, D. D., n.d. *Schooling, Labor-Force Quality, and the Growth of Nations,* s.l.: s.n.
6. Doucouliagos, H., & Abdullah, A. (2013). *Does education reduce income inequality? A meta-regression analysis*. Journal of Economic Surveys, 29(2), 301–316.