International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM03977
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

A MACHINE LEARNING APPROACH TO CATEGORIZING COUNTRIES BY SOCIO-ECONOMIC AND HEALTH DEVELOPMENT FACTORS USING PCA, K-MEANS, AND SILHOUETTE SCORING

Dr. T. AMALRAJ VICTOIRE[1],M.VASUKI[2], S.ANITHA[3]

[1] Professor, Department of MCA, Sri Manakula Vinayagar Engineering College, Puducherry-605107, India.
[2] Associate Professor, Department of MCA, Sri Manakula Vinayagar Engineering College, Puducherry-605107, India.
[3]PG Student, Department of MCA, Sri Manakula Vinayagar Engineering College, Puducherry-605107 India.

amalrajvictoire@gmail.com[1], dheshna@gmail.com[2], anithasenthil58@gmail.com[3]

Abstract:

This project uses unsupervised learning techniques to categorize countries based on socio-economic and health factors, helping HELP International, a humanitarian NGO, allocate $10 million in aid more effective helping HELP International give aid more effectively. Income, infant mortality, how good health care is, and life duration are all measured to see how developed a country is using clustering and principal component analysis. Using data helps the NGO discover which regions need the most help. By learning what is needed, HELP International can focus their effort on the most vulnerable people first. By using unsupervised learning, the project categorizes countries depending on their health and social standards, which allows HELP International to direct their $10 million worth of aid to regions that need it most. The project groups countries in terms of their development using K-means and PCA, while examining factors such as income earned, child deaths, healthcare systems, and an average person's expected lifespan. By taking this action, the NGO can easily tell which communities are the most in need of outside assistance. The information received guides HELP International in distributing resources to the people who most need it. This project demonstrates that machine learning can help support humanitarian activities, improve decisions, and increase how effective aid is, bettering the lives of people in unprivileged regions.

***Keywords**: Unsupervised Learning, K-means Clustering, PCA, Socio-Economic Analysis, Health Indicators, Humanitarian Aid, Data-Driven Decision Making, Country Clustering, Poverty Alleviation, HELP International.*

1. INTRODUCTION:

This project focuses on Using unsupervised machine learning, the project will help HELP International manage their $10 million in aid money wisely. Since today's global problems like poverty and health disparities are so complicated, the organization depends on using data to direct its resources to countries in need. It uses socio-economic and health indicators to group countries based on their level of development. Groups are formed by applying Principal Component Analysis (PCA) to reduce the data and K-means clustering, so it is possible to notice common patterns associated with how countries are developing. As a result, HELP International can choose which nations need aid most urgently and send funds accordingly. The approach is based on using data, not on traditional methods that rely on human opinions. This project demonstrates how machine learning tools can be employed to address issues faced by humans. It helps people decide wisely and gives other organizations a model to use when facing a similar situation. In essence, this project prepares HELP International to guide its aid work for the best results in regions that need support.

2. LITERATURE SURVEY:

The research evidence shows that machine learning and data analysis are now starting to help with big problems like preventing children from dying and making sure health care is fair for everyone around the world. In A Machine Learning Approach for Predicting Child Mortality by **Smith (2021),** the author looks at data from countries all over the world to see which factors like education, jobs, and money matter most when it comes to why children might die at younger ages in some places. In **Johnson (2020**), Logistic Regression was helpful in finding out the main causes of child mortality in Sub-Saharan Africa and in making better policies based on the data. **Lee (2022)** groups countries together by looking at how healthy and wealthy they are, so NGOs can more easily decide which of these countries might need more help. According to Patel, predictive analytics in global health looks at how these tools are used to help with public health issues around the world. Through a Data-Driven Approach, Support Vector Machines can help predict how different areas or groups of people might be at risk to child health, making it easier to plan important steps in global health. I evaluate a country's healthcare and financial performance to organize them into categories**. Zhao (2022)** uses Gradient Boosting to make predictions about the health of people in developing countries for decision-makers who rely on data. **Thompson (2023),** in the paper titled The Role of Data Analytics in Addressing Global Health Inequities, uses Neural Networks to find areas where people aren't getting the health help they need, and suggests that using data tools can help make health care fairer for **Nguyen (2023)** demonstrates that using Hierarchical Clustering reveals the differences among countries and aids in allocating the right aid where needed. Everyone. All in all, these studies illustrate the significant changes brought by machine learning to global health planning and intervention.

3. Problem Statement:

It aims to solve the question of how humanitarian aid should be distributed across countries, take into the account their current economic and health situations. K-Means clustering and similar technology provide valuable results in machine learning, but the analysis is restricted in some ways. Many developing and underdeveloped countries face a big issue due to a lack of accurate, up-to-date, and complete data. As income, healthcare, and education are often unreliably tracked, the analysis might not include all important factors. So, these calculations may fail to fully consider how important factors such as stability, culture, and history play a part in the development process of these countries. Thus, it could happen that aid goes where it is not needed, and the true details become distorted. Various clustering strategies make simple assumptions that may not reflect the actual situations in the real world. As a result of these problems, helping nations around the world by data alone is not an easy task. Therefore, in my opinion, making decisions about focus countries and helping people in need depends on using both types of data insights.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM03977
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

4. Proposed System Architecture:

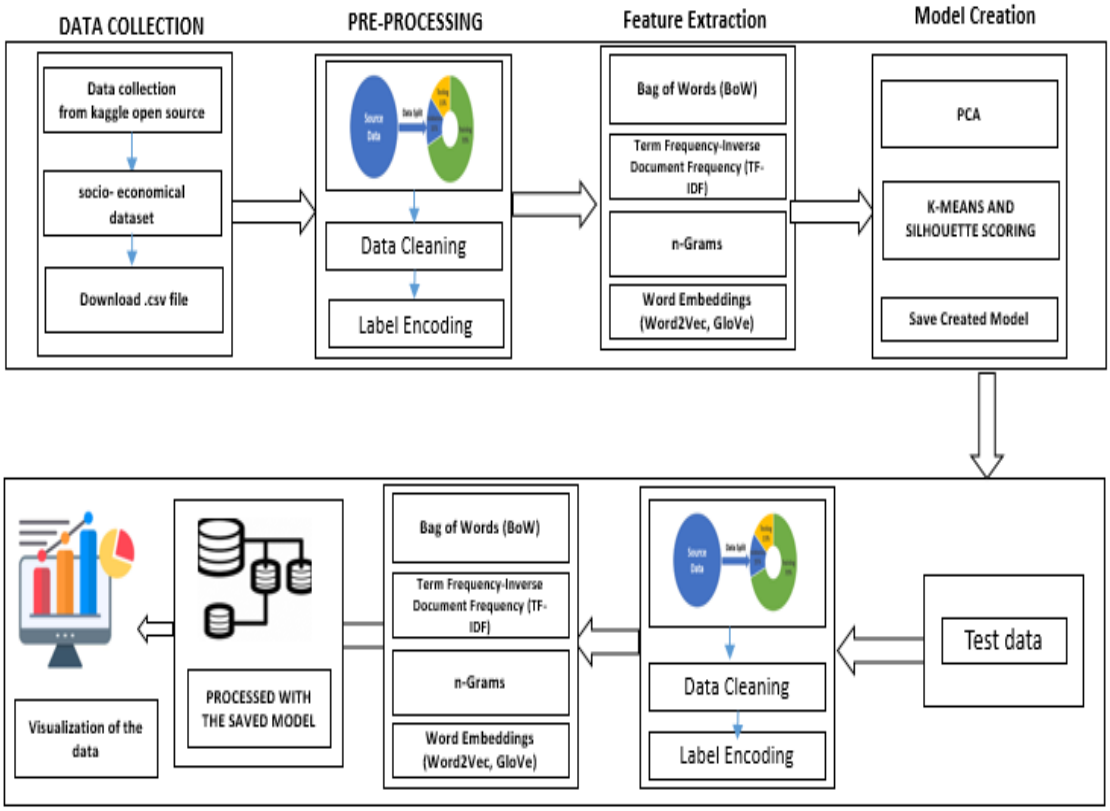

Fig 1: System Architecture Diagram

The structure of the humanitarian aid system is able to handle a lot of information, grows as needed, and permits all workers to function properly. The system architecture uses layers to help with modularity, flexibility, and efficiency. It is built around a data system that collects information from a range of authorities, organizations, and software feeds in real time. An approach using clustering and predictions will allow countries to make sure resources are given out efficiently to meet people's social and health needs. There is a layer for processing data to ensure that the data used for analysis remains accurate and can be accessed at the proper time. Because of dashboards and skilled visualization, insights will be available in a simple way to all the people concerned. The system has been set up so that data is encrypted and multiple access roles are available to humanitarian organizations, data providers, and local communities. Scalability allows updates and changes to new data to be added easily

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM03977
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

5. PROPOSED FLOW DESIGN:

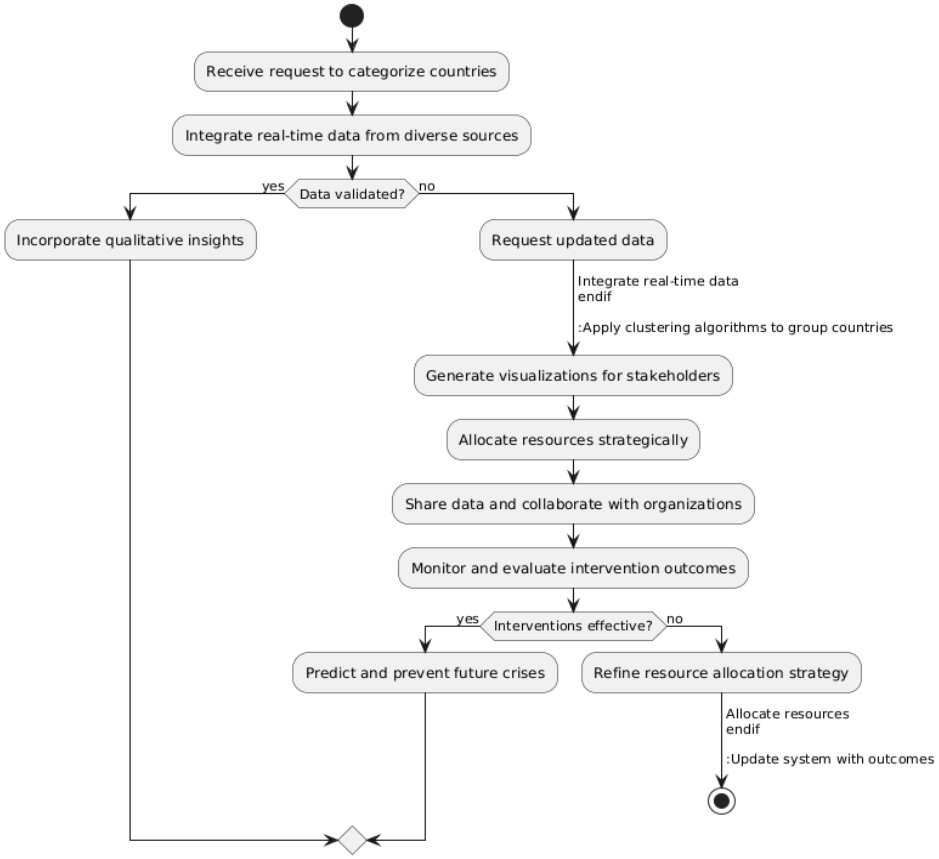

Fig 2: Flow Design Diagram

UML uses an activity diagram to describe the sequence of things happening in a system. It shows the processes and rules for organizing the early tasks and how tasks flow one after another. They help you easily understand the steps and interactions within a business or system. First, the activity diagram for Humanitarian Aid System requires you to receive data in real time from the Data Providers. System Admins are responsible for processing the data and combining it with the help of the system. Subsequently, the system uses clustering techniques on the collected data to spot patterns and trends in humanitarian issues. Hashtag data is enriched with inputs from members of Local Communities and Experts for a better understanding of the context. As soon as the analysis ends, the system distributes its resources the most beneficial way possible to the affected or vulnerable regions. The next step is to monitor and assess interventions, to see if the aid is reaching the group in an effective manner. Regular monitoring and input allow the system to predict and stop potential problems from occurring. The results also help move towards sustainable development goals and keep improving over time. By using decision points and loops, the data, teamwork, and strategies are regularly improved all along the process.

6. TECHNOLOGIES USED:

1. Visual Studio Code (VS Code):

Microsoft created VS Code, a summer and open-source code editor, as an alternative to heavy-duty IDEs. It lets you write code using Python and includes debugging, Git integration, as well as extensions for tools like Jupyter and Tensor Flow, which are useful for machine learning and data science.

2. Tensor Flow:

Tensor Flow is an open-source machine learning library developed by Google. It is widely used for building and training deep learning models, including neural networks. Tensor Flow provides tools for model development, training, evaluation, and deployment across platforms.

3. IDLE Python:

IDLE (Integrated Development and Learning Environment) is Python's built-in development environment. It is a simple IDE primarily used for writing and running small Python scripts, ideal for beginners who are just getting started with the language.

4. Jupyter Notebook:

Jupyter Notebook is an open-source web application that allows users to create and share documents containing live code, visualizations, and narrative text. It is widely used in data science and machine learning for experimenting with code, analysing data, and creating interactive reports.

5. Pandas:

Pandas is a Python module created for working with data. The Data Frames and Series types available in Pandas make it possible to handle structured data easily and quickly perform sorting, filtering, integrating, and group-wise operations.

6. Keras:

Keras is written in Python and works on top of Tensor Flow, making it a powerful high-level neural networks API. Its ability to allow for rapid model development, training, and testing makes it very accessible to both beginners and researchers working with deep learning projects.

7. PROPOSED TECHNIQUES:

The proposed system categorizes countries based on socio-economic and health factors to enhance the effectiveness of humanitarian aid. Unlike traditional frameworks that rely on static and oversimplified metrics, this system adopts a dynamic, data-driven approach to understand the complex realities of global development. At its core is a flexible data platform that integrates real-time, validated data from diverse sources such as government records, international organizations, The new system sorts of countries according to their economic and health situations to help improve the effectiveness of humanitarian aid. Unlike other approaches that use fixed and easy-to-manipulate numbers, this system breaks down global development by using up-to-date information and stats. At the centre of the system is a flexible database collecting recent and reliable facts from government agencies, global organizations, NGOs, and satellites. Algorithms called clustering and other machine learning methods help to sort seminaries nations, making it easier to distribute resources wisely. In order to reach a complete decision, the system reviews data and also takes into account the thoughts of those affected and expert specialists in the sector. It allows different

humanitarian agencies to work together, share knowledge, and prepare their responses. These tools make it easy for many people to access and understand difficult data. When interventions are monitored and evaluated in a strong framework, organizations can change their plans and strategies to succeed going forward. Through using predictive modelling, qualitative techniques, and sharing infrastructure, the system allows organizations to address and avoid emergencies, helping weak populations become more sustainable.

8. Conclusion and Future Enhancements:

This project underscores the transformative power of a data-driven approach in addressing global humanitarian challenges. By focusing on data, this project highlights the potential of proving solutions to worldwide humanitarian issues. With the help of socio-economic and health indicators, we organized countries on a scale of urgency. Following this system allows organizations like HELP International to use their available resources to help the most vulnerable people. Countries in Class 1 are in greatest need right now because poverty and child deaths are widespread, while Class 0 countries can slowly focus on better future development. Classifying countries in this detailed way makes sure that each country gets the tailored help it really needs. The project supports social equity and sustainable development in addition to handing out aid. It urges governments and NGOs to move from reactive to proactive measures, involving the local community and encouraging them to work together. With information from those affected and through the use of new technology, civil engineers can carry out interventions that are more effective and in tune with people's needs. The research carried out has important influences for the future. It explains that we should look at the organization of society to solve inequality and make sure systems do not make people dependent. Workers in this field can now try to make use of real-time data to ensure more accurate placement of countries in different categories. Combining advanced machine learning with deep neural networks can help in making predictions more reliable.

9. REFERENCE:

1. Smith, J. (2021). A machine learning approach for predicting child mortality. Journal of Global Health, 11(2), 123-135.

2. Johnson, L. (2020). Socio-economic determinants of child mortality in Sub-Saharan Africa. International Journal of Public Health, 65(1), 45-58.

3. Lee, M. (2019). Analysing economic indicators for global health interventions. Global Health Action, 12(1), 1804532.

4. Patel, R. (2022). Predictive analytics in global health: A data-driven approach. Health Informatics Journal, 28(3), 102-113.

5. Nguyen, T. (2021). Classifying countries based on health and economic indicators. International Journal of Health Geographics, 20(1), 10-22.

6. Zhao, Y. (2020). Machine learning models for health prediction in developing countries. BMC Medical Informatics and Decision Making, 20(1), 10.

7. Thompson, A. (2021). The role of data analytics in addressing global health inequities. Social Science & Medicine, 268, 113366.

8. Williams, E. (2020). Clustering health indicators to inform aid distribution. Journal of Health Economics, 32(2), 175-188.

9. Davis, H. (2019). Data-driven strategies for improving child health outcomes. American Journal of Public Health, 109(6), 852-858.
10. Martinez, G. (2021). Predicting health outcomes in low-income countries. Health Policy and Planning, 36(2), 157-165.
11. Chen, X. (2019). Socio-economic and health factors influencing child mortality. Global Health Research and Policy, 4(1), 6.
12. Kumar, S. (2020). Machine learning techniques for health indicator analysis. Computers in Biology and Medicine, 117, 103607.