

## **Capstone Project Proposal**

### **Problem**

Donors are concerned with practical allocation of aid towards developing countries. The elimination of poverty worldwide is the first of seventeen goals for 2030<sup>1</sup>, but it requires frequent, reliable data on poverty indicators and distribution. Measuring poverty is critical to eliminating it<sup>2</sup>.

Apart from the limited economic data of developing countries, there are many practical questions that donor country officials face in terms of prioritization of aid allocation: will aid help the country if the fragility is a result of poor governance? Should allocation be based primary on income levels, or should the donor contribute to countries with struggling economies instead? What factors contribute to the poverty in a particular country and what aspects should the aid specifically target to rectify these foundational issues?<sup>3</sup>

Researchers need more data and a better way to verify income qualification (currently, the Proxy Means Test<sup>4</sup> and the World Development Indicators<sup>5</sup> are used). By using the data we currently have from developing countries, we can see what factors, or features, best predict poverty. This can narrow down the type of data researchers need to collect and help donors focus on specific areas to allocate aid to (such as education, housing, government assistance). Identifying poverty, the regions most in need<sup>6</sup>, and the infrastructure that is key to poverty elimination are key components to eliminating world poverty.

### **Data + Approach**

The Inter-American Development Bank has data for Costa Rican household characteristics, which it has collected in an effort to explore new methods beyond traditional econometrics for evaluating social need.

In this project, this data will be analyzed to understand what factors (which can be perhaps be more easily collected in developing countries than traditional econometrics) most contribute to and can best predict poverty. The information from the final model can be used to extrapolate what specific data should be collected in countries that have a similar profile to Costa Rica.

The data is available through [Kaggle](#).

### **Product**

The final deliverable will be a web application that will allow a user to enter information for relevant features that the model uses. The output of the web application will be an income level indicator, which predicts the poverty level based on the input fields, and information regarding the level of social aid required.

Machine learning classification algorithms will be used to build the model. Random forests and XGBoost will be applied first.

Deep learning will also be used in order to see what benefits or challenges come from the feature extraction component of deep learning algorithms. The machine learning and deep learning techniques will be compared, and the final product will include the technique that provides the best accuracy.

### **Requirements**

This project will be completed using Domino and Paperspace, as provided by the Springboard course. Currently, no additional CPUs or GPUs are required in order to build the machine and deep learning models.

### **Conclusion**

By analyzing the Costa Rican household dataset, researchers will be able to determine what factors contribute most to income level predictions, and can help donors by targeting aid to specific regions and towards specific foundational issues.

### **Footnotes**

1. <http://sustain.stanford.edu/predicting-poverty>
2. <https://www.drivendata.org/competitions/50/worldbank-poverty-prediction/page/98/>
3. <https://www.brookings.edu/research/how-should-official-development-assistance-be-allocated-across-countries/>
4. <https://www.kaggle.com/c/costa-rican-household-poverty-prediction/overview>
5. <https://datatopics.worldbank.org/world-development-indicators/themes/poverty-and-inequality.html>
6. <https://bigcloud.io/how-ai-can-help-alleviate-poverty/>