

MODELING, VISUALIZATION, AND ANALYSIS OF AFRICAN INNOVATION PERFORMANCE

Anonymous authors

Paper under double-blind review

ABSTRACT

In this paper we discuss the concepts and emergence of Innovation Performance, and how to quantify it, primarily working with data from the Global Innovation Index, with emphasis on the African Innovation Performance. We briefly overview existing literature on using machine learning for modeling innovation performance, and use simple machine learning techniques, to analyze and predict the "Mobile App Creation Indicator" from the Global Innovation Index, by using insights from the stack-overflow developers survey. Also, we build and compare models to predict the Innovation Output Sub-index, also from the Global Innovation Index.

1 INTRODUCTION

Measuring innovation of a certain nation or region emerged as a common field of study in the few last years; as the levels of complexity of each nation and the existing policy structure and infrastructure prove a major challenge in creating a standardized system that can be generalized over regions or over the globe. National innovation performance measures do exist in part of the world to pass policies and to explore and come up with improvements in the infrastructure and policy towards innovation in that country. One was the report created by the Advisory Committee of Measuring Innovation in the 21st century economy and presented to the US Secretary of Commerce in 2008, which – among-st other goals – aimed to develop better ways to quantify innovation in the marketplace and to guide the government towards creating frameworks for measuring innovation and direct the policies that aim to uplift innovation performance in the united states. Work on regional innovation metrics was a bit more successful. The European Union publishes a yearly report that summarizes the innovative performance of the region called "The Regional Innovation Scoreboard". This scoreboard divided the EU regions into four distinct classes, Regional Innovation Leaders, Regional Strong Innovators, Regional Moderate Innovators, and Regional Modest Innovators. Finding a global unified index is an especially difficult task. Where for example pioneering work by Michael E. Porter et al; on the National Innovative Capacity provides a good historical overview on the countries it covers; the diversity of these countries is the problem as it only takes a look at a specific sector of the world mainly the U.S, Europe and the high to upper-middle income Asian countries with no representation of third world countries and the African region to be exact. The Global Innovation Index, which we have chosen as a benchmark for our analysis in this paper, helps to create an environment in which innovation factors are continually evaluated. It is divided into two major indices; the Innovation Input sub-index and the Innovation output sub-index. The innovation input sub-index has five factor evaluating institutions, Human Capital and research, Infrastructure, Market Sophistication and Business sophistication while the Innovation output sub-index gives insights on knowledge and technology outputs and the creative output. The GII solves the problem of representation as it covers 129 countries from different income classes in its 2019 report. However, the problem for quantifying innovation performance worldwide, and in particular, Africa, is far from solved. The Global Innovation Index is mutilated here and there by missing data and – arguably – misinforming indicators. Where most of country profiles incomplete come from developing countries specially Africa, we take on one indicator – Mobile App Creation – that has a high ratio of missing data from the observations in the very same 2019 report, and try to build a robust estimator to interpolate over its values, as we will discuss in the next section. Finally, employing machine learning techniques in the analysis and modeling of such innovation performance measures is a promising approach with respect to building robust predictors and providing deep insights that

might lead to novel ideas. That being said, however, contributions from the machine learning community with respect to this topic are very few, not to mention that none of them addresses Africa in particular. Two works in particular we think should be addressed are by Forner, et al; and Hajek & Henriques, respectively. The first, which we follow a similar approach to it here, performs their analysis on the Global Innovation Index, and the second multi-output artificial neural networks to model regional European innovation, operating on data from the EU's "Nomenclature of Territorial Units for Statistics". Both of these works conclude with the remark that machine learning algorithms perform better on modeling innovation data than traditional analytical methods popular in the literature. .

2 METHODOLOGY

Our two main contributions in this paper are building a model that uses insights from alternative data sources to predict the Mobile App Creation indicator in the Global Innovation Index worldwide data for 2019, and building a model to predict the Innovation Output Sub-Index for a set African countries over the last six years (2014 - 2019), also using the Global Innovation Index data for the indicated years. For the first one, we used survey data provided by stack-overflow in their website. The survey is organized by stack-overflow, and over 90,000 developers participate in the survey each year, it provides for us deep and comprehensive insights on the state of software development in every country in the world. The survey, which in the form of multiple choice questions, has more than 300 questions and takes about 20 minutes to complete. We have chosen a set of 30 questions from this survey, we believe best represent the status quo of each country's local software market, as well as core competencies for each developer along with their corresponding countries. We then removed unique IDs and averaged one hot encoded column values over the countries, to produce a structure similar to the Global Innovation Index', and finally merged the data set with the Mobile App Creation indicator data, with respect to countries. After that, a correlation matrix was produced to learn about the interaction between features, and determine whether the developer survey's data was relevant at all to the Mobile App Creation Indicator. Finally, we built four models; a Gaussian Process model, an extremely Gradient Boosted Trees (XGBoost) model, Support Vector Machine (SVM) model, and a Random Forests model, and compared their performance using using k-fold cross-validation with $k = 10$, and with the root mean squared error (RMSE) as a measure. The models accepted all the averaged/weighted survey questions and predicted the Mobile App Creation Indicator. For the second contribution, we followed an approach similar to the one employed in Forner, et al. We used the extended report which has 81 detailed features of the 7 aforementioned indices. We extracted the data for Africa for the last six years spanning from 2014 until 2019 and covering 36 African countries with some only being represented in 2 years. Unlike Forner et al; we didn't drop out countries with incomplete profiles, since that would only end up in eliminating half of the African countries from the data set, also, our analysis focuses on only African countries, with keen consideration to the context and aspects of the shortcomings in developing countries, rather than merely building a predictive model for global innovation performance. We as well, produced a correlation matrix, and employed the four models mentioned above in this analysis as well, the only difference being that this time, the models accepts the 81 features described, and outputs a prediction of the innovation output sub-index. Finally, we visualized the fluctuation in innovation performance through these years on a geographical map, using geopandas.

3 RESULTS

For the Mobile App Creation prediction, it was found that the data collected from the surveys as weakly correlated with the indicator, which has resulted in the weak performance of the models, except for the Gaussian Processes model which performed best performed best. These poor results, come from the fact that the Global Innovation Index employs a top-bottom approach in the way it collects its data to compute the indicators. For the Mobile App Creation case in particular, it uses "Global downloads of mobile apps, by origin of the HQ firm, scaled by PPP\$ DGP (billions)". This approach explicitly eliminates local innovators, specially those in developing countries who mostly exist as freelancers, social enterprises, or early stage start-ups. This calls for a grassroots movement that extends beyond merely assuming better quantifiers for innovation performance, as well as the

pressing need to use alternative data sources to have accurate estimates for quantifying Innovation Performance in the future, specially in the developing world.

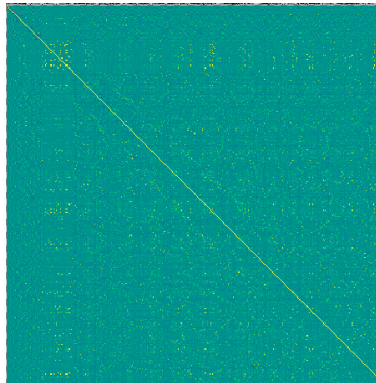


Figure 1: Correlation Matrix for Developer's Survey

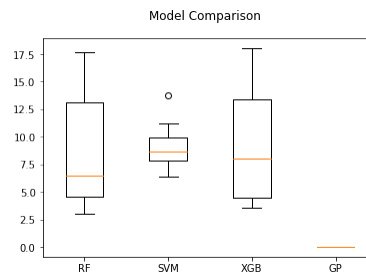


Figure 2: Model Comparison on Developer's Survey Data

As for the African countries data set, output of the correlation matrix showed that there is a direct positive correlation between Regulatory Quality of a country and its Innovation output sub-index, meaning an active role of government can lead to a healthier innovative environment and therefore better innovation overall. Another positive correlation appeared between the Innovation Output sub index and Governments online service and also with Rule of law continuing the apparently needed state sponsoring of innovative friendly policies for acceleration to show in innovation. One of the most interesting correlation was that ISO 14001, an environment certificate appeared in very high correlation with the Innovation output which means that a moving towards a clean environment could push Africa's innovation performance further. Finally the models performed relatively very good, with XGBoost being the dominant with no surprise.

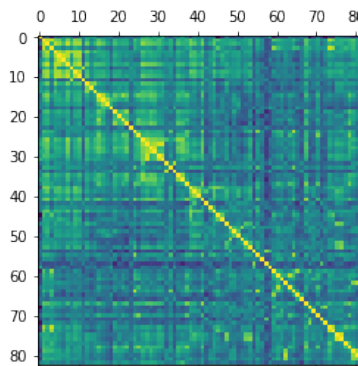


Figure 3: Correlation Matrix for Global Innovation Index Indicators

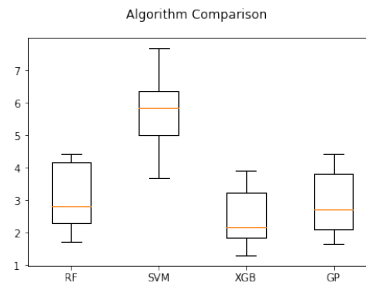


Figure 4: Model Comparison on Global Innovation Index Indicators

4 CONCLUSION

In this paper, we outlined the need to use alternative data sources to better measure innovation performance in the continent, as well as the need for grassroots movements to foster and facilitate innovation in Africa. We also showed how simple machine learning techniques can provide novel insights for this matter in question, and finally, a number of observations considering African innovation performance and recommendations to better facilitate innovation and creativity in Africa.

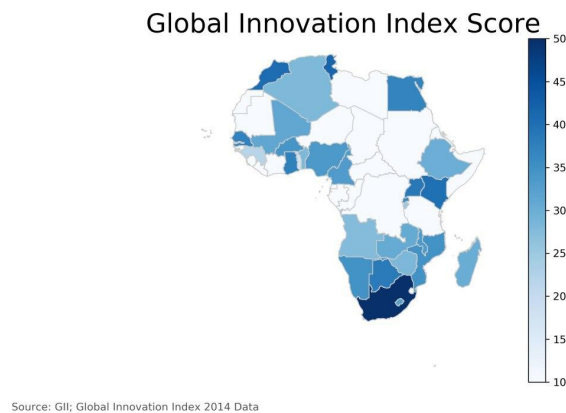


Figure 5: African Innovation Performance in 2014

5 REFERENCES

Hadzimustafa, Shenaj & Rexhepi, Gadaf. (2011). Measuring Innovation in the 21st Century Economy. SSRN Electronic Journal. 10.2139/ssrn.1929039

Global Innovation Index. 2020. Global Innovation Index — Creating Healthy Lives-The Future Of Medical Innovation. [online] Available at: ;<https://www.globalinnovationindex.org/Home>; [Accessed 11 March 2020].

Furman, Jeffrey & Porter, Michael & Stern, Scott. (2002). The Determinants of National Innovative Capacity. Research Policy. 31. 899-933. 10.1016/S0048-7333(01)00152-4.

Global Innovation Index. 2020. Global Innovation Index — Creating Healthy Lives-The Future Of Medical Innovation. [online] Available at: ;<https://www.globalinnovationindex.org/Home>; [Accessed 11 March 2020].

Fornier, Dominik & Ozcan, Sercan & Bacon, David. (2019). Machine Learning Approach for National Innovation Performance Data Analysis. 325-331. 10.5220/0007953603250331.

Hjek, Petr & Henriques, Roberto. (2017). Modelling innovation performance of European regions using multi-output neural networks. PLOS ONE. 12. e0185755. 10.1371/journal.pone.0185755.