

Project Proposal (Artificial Incoherence)

Members:

Francisco Manuel Mares Solano
Trang Tran
Wenbin Jiang
Xindi Zheng

1) Problem

Lack of opportunities, discrimination, crime, and other problems that affect societies worldwide have gained significant attention among developed countries, where the topics are discussed and dissected across different criteria, such as race, gender, religion, economical status, etc. Although the investment in researching this topic and its solutions remains on-going, these countries have taken a very important step towards equality and justice: awareness.

In contrast, so-called 'underdeveloped' countries often deal with these issues in silence, as they usually face other (and equally important) problems, and thus, the situation becomes somehow normalized, taboo, or unapproachable due to a lack of resources.

However, in the case of Mexico, the problem has not been the lack of resources. Proof of that is the national census which registers the disparities in its society every few years. The problem is the enormous socio-economic gap which separates the upper class from the lower class into bubbles that are completely isolated from each other, and within which the middle class is almost non-existent. In Mexico, the social class with the resources to research the topics related to discrimination does not believe that inequality exists in general terms or that it is an issue at all, while the lower class that suffers discrimination simply lacks the resources to solve it.

In this project, we will delve into the available data sources, to try to provide a panorama on the impact that the different socio-economic factors can have among Mexican citizens. In doing so, we will gather, analyze and compare different data sources to attempt to detect the patterns that describe the social tendencies that have been observed in many other countries. We expect that this research will aid in cementing the bases for a formal analysis of the inequality in the country.

2) Data mining task

Clustering and classification

3) Data

In official journals, reports and governmental census

4) Schedule

week 1 & 2:

Data collection.

We will investigate the different available data sources to have the information as complete as possible

week 3 :

Data exploration.

We will analyze the different data sources, we will clean them, normalize them, and will find correlation patterns to create an argument for our research.

week 4,5.

Experimentation.

We will test different approaches, such as different algorithms, different configurations, different visualization methods, etc., to try to determine the most effective methods.

week 6,7.

Improvement

Having determined the most effective approach, we will further experiment with its parameter to try to better approximate our model.

week 8,9.

Analysis of results

We will analyze our results to try to provide valuable conclusions, as well as produce a systematic report of our approach, limitations and observations.