MANAGEMENT SUMMARY GROUP 2

THE LEVELS OF CORRUPTION

PATRICK FRENZEL, SHRISTI RAJBAMSHI, CARLOTTA TUBEUF, JOANNA ZAMIECHOWSKA

How can the level of corruption of a country be measured?

In general, corruption can be described as the abuse of a given position of trust for private gain. However, there is no universal definition of the term, let alone an objective measurement of a country's level of corruption. The Transparency International movement's mission is to "stop corruption and promote transparency, accountability and integrity at all levels and across all sectors of society[1]". To achieve this transparency, every year, they evaluate the Corruption Perceptions Index (CPI) for each country based on how corrupt their public sectors are perceived to be. The CPI ranges from 0 to 100, whereby a small value indicates a high corruption rate and vice versa. Because this Index considers numerous factors, is renowned worldwide and has an extensive and transparent data set, we chose it as our primary source for measuring a country's degree of corruption. Figure 1 shows the percentual change in CPI from 2000 to 2020 for all countries.

Considering that the other databases that we explored looking for different ways to measure the level of corruption (mainly The World Bank - Worldwide Governance Indicators[2]) didn't cover enough countries or years in order to gain valuable insights, we focused on the Transparency International CPI Score and on the question of which other characteristics of a country are influencing the level corruption.

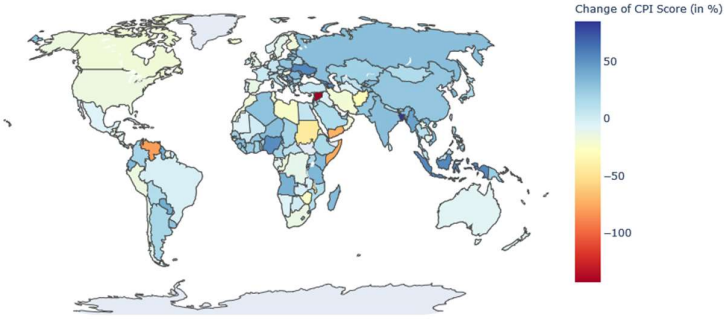

Figure 1: Percentage difference of CPI between 2000 and 2001

What characteristics of a country predict the level of corruption?

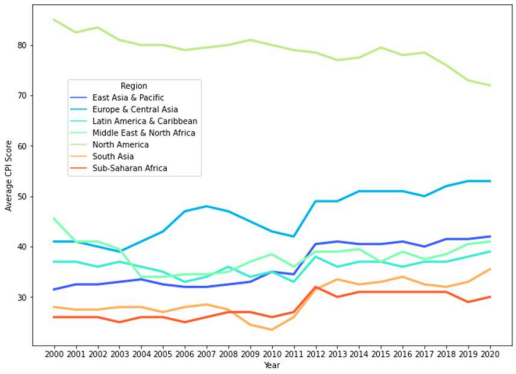

Figure 2: Average CPI over years for different regions

By dividing the countries into regions and income groups, first assumptions on which characteristics may correlate with the level of corruption can be made. Looking at the average CPI score distribution over time in Figure 2, it can be assumed that highly developed western countries are less corrupt than countries in the global south. The overall trend for the last 20 years is that corruption decreased worldwide except for North America, where the perceived level of corruption (more or less) constantly diminished. Figure 3 visualizes the distribution of the CPI over the income groups. It is evident that the CPI score increases with the

[1] https://www.transparency.org/en/about, last accessed 2022-01-23
[2] https://databank.worldbank.org/home, last accessed 2022-01-23

augmentation of a country's wealth. Hence, countries with lower average income are perceived as more corrupt than countries in higher income groups.

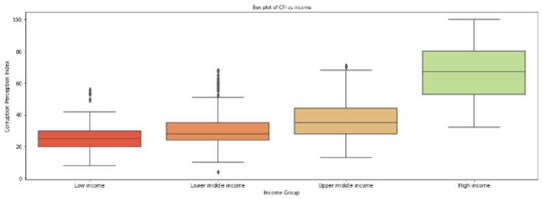

Figure 3: Boxplots for average CPI over income groups

After literature and internet research on existing assumptions for characteristics that correlate with corruption, we established hypotheses for predicting a country's level of corruption following an iterative process. We tested our assumptions by calculating the correlations between the characteristics and the CPI and added more data of possibly correlating attributes if no strong linear correlations were detected. The final characteristics based on which we tried to predict the level of corruption of a country (in addition to the region and the income group) were:

- GDP
- Gender Equality
- Expenditure on Education
- Political Stability
- Regulatory Quality
- Unemployment
- Population Size
- Ease of Doing Business
- Renewable Energy Consumption
- Population Growth
- Population Density
- Expenditure on Military
- Land Area
- Alcohol Consumption
- Life Expectancy
- Inflation
- Infant Mortality
- Suicide Rate

All the data for the characteristics listed above was downloaded from the World Bank data set[2], from several databases (see the corresponding python notebook for more detailed information on the sources). Based on these characteristics, we tried to predict the CPI score using several machine learning algorithms. Figure 4 shows the comparison of the results from the predictions with linear regression, decision tree and kNN. Using kNN, we were able to determine the CPI score for a given country with an accuracy of 92% based on the above characteristics.

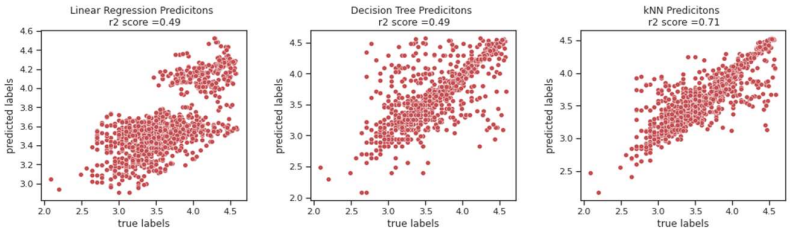

Figure 4: CPI prediction results for linear regression, decision tree and kNN

Conclusions

The answer to the question of how a change in the level of corruption of a given country can be predicted was not found as straightforward as we initially thought it would be. The characteristics that we looked into at first seemed to not correlate as strongly as anticipated. This shows that the measurement of corruption is a highly complex matter that can only be confirmed by the lack of a universal definition and quantification for corruption. Nevertheless, we succeeded in predicting the CPI score with an accuracy of over 90% based on various correlating attributes. The most significant takeaways are:

1. Since 2000, the **CPI score decreased** for most parts of the world.
2. With **increasing wealth**, the level of corruption of a country decreases.
3. If a country's **regulatory performance** regarding the ease of doing business is good, corruption is likely to decrease over time.

Overall, it should be noted that bias could have been introduced in our selection of attributes as these are attributes that we, personally, tend to associate with corruption. Bias could also be present in the partly self-reported CPI scores. For future work it would be interesting to look into further possibly correlating characteristics such as nepotistic appointment of government officials, government type, primary economy types, main exports, etc.