Globalization in Data Science Education: A Case Study

Hossein Saiedian
Electrical Engineering and Computer Science
The University of Kansas

Abstract

The increasing interconnectedness of our world demands a global perspective in data science education. By analyzing international data on poverty, health, national debt, and education, students gained essential data science skills and insights into global disparities and trends. The project provided hands-on experience with data acquisition, exploratory analysis, predictive modeling, clustering, and inferential analysis using datasets from the United Nations, World Bank, and IMF. Students developed practical skills in data wrangling, visualization, and modeling, while gaining a deeper understanding of global issues. Through this project, students enhanced their critical thinking, problem-solving abilities, and global awareness. The case study demonstrates the value of incorporating globalization into data science education to prepare students to address complex global challenges and contribute to a more equitable and sustainable world.

Keywords. Globalization, data science education, socioeconomic analysis, predictive modeling, statistical inference

Introduction

We present a case study of a "Global Socioeconomic Analysis" project integrated into an introductory data science course at the University of Kansas. The project aimed to equip students with essential data science skills while fostering a global mindset. By analyzing international data on poverty, health, national debt, and education, students gained insights into global disparities and trends. This paper demonstrates the value of globalization in data science education and provides a framework for incorporating international perspectives into data science curricula.

The "Global Socioeconomic Analysis" project was designed to provide students with hands-on experience applying data science techniques to real-world global challenges. Students worked with datasets from international organizations such as the United Nations, World Bank, and International Monetary Fund. The project encompassed the following stages:

- **Data acquisition and preparation:** Students collected, cleaned, and integrated diverse datasets, addressing issues such as missing values and inconsistencies.
- **Exploratory data analysis:** Using visualization techniques, students explored data distributions, identified patterns, and uncovered relationships between variables.
- **Predictive modeling:** Students built models to predict socioeconomic outcomes based on various factors, such as poverty rates, life expectancy, and GDP per capita.
- **Clustering analysis:** By grouping countries with similar characteristics, students identified socioeconomic clusters and analyzed their implications.
- **Inferential analysis:** Students tested hypotheses about global trends and drew conclusions based on statistical evidence.

Why Globalization

Opens students' data science horizons. International data unlocks a diverse spectrum of perspectives, challenges, and opportunities, enhancing practical skills in Python programming and data science methods. The exposure to real-world datasets from various countries provides a hands-on opportunity to apply Python programming skills to address complex global issues, fostering the development of practical, industry-relevant expertise.

Helps students embrace the interconnected world. In today's interconnected world, understanding data across borders is crucial. The students skills can be used to tackle global issues like climate change, poverty, and disease, making a student a valuable part of the international problem-solving team. This involvement facilitates the development of a global perspective on socioeconomic issues, as they gain insights into the interconnected nature of global challenges and their socioeconomic implications.

Boosts students career potential. Fluency in international data sets is a sought-after skill in a globalized tech market. By demonstrating their cross-cultural data analysis capabilities, students position themselves as a valuable asset to organizations with international interests. This experience enhances students employability and opens doors to diverse career opportunities in the ever-expanding field of data science, concurrently providing practical experience in working with real-world, large-scale datasets.

Global impact, local insights. Students will explore data from diverse corners of the world to reveal hidden patterns and trends invisible within national borders. Their analysis can inform solutions to global challenges like climate change or resource scarcity, making a real-world impact beyond classroom. This benefit not only exposes students to large-scale datasets but also allows them to develop a global perspective on socioeconomic issues, contributing to a deeper understanding of the impact of data analysis on global problem-solving.

Project Objectives and Scope

The "Global Socioeconomic Analysis" project is designed to provide students with hands-on experience in applying data science techniques to real-world global issues. The primary objectives of the project are as follows.

Data Wrangling and Preparation: Students in this data science project will gather data from various international sources, including institutions such as the United Nations, International Monetary Fund, World Bank, and Eurostat. After collecting this diverse dataset, they will meticulously cleanse and preprocess it. Their tasks involve handling missing values, addressing outliers, and ensuring overall data accuracy and consistency. The result will be a structured and ready-to-use comprehensive dataset for subsequent analysis.

Visualization: As part of their project, students will create visualizations to explore patterns, trends, and relationships within the datasets. These visual tools may include histograms, scatter plots, heatmaps, and other graphical representations. Additionally, they will delve into dynamic and interactive charts to enhance their exploratory data analysis process. Geographic visualizations will also play a role in showcasing global patterns and highlighting differences across regions.

Classification and Regression: In this stage, students will build predictive models using classification and regression techniques. Their goal is to identify factors influencing socioeconomic indicators. For instance, they might investigate how poverty rates, health outcomes, or national debt impact specific metrics. To ensure robust models, they will evaluate key features and performance metrics rigorously.

Clustering: Students will apply clustering techniques to identify distinct socioeconomic clusters or groups within the global datasets. Algorithms will help them group similar data points together based on certain characteristics. As part of their analysis, they'll explore whether these clusters align with geographical or economic factors. Furthermore, they'll assess the implications of these clusters on global socioeconomic dynamics.

Statistical Inference: Formulating inferential questions based on associations observed during exploratory analysis is a critical step. Students will then apply statistical tests to validate these associations or draw meaningful conclusions from their data.

Remember that hands-on experience throughout these stages is crucial for becoming proficient in data science!

Key Findings and Insights

Through this project, students gained a comprehensive understanding of global socioeconomic disparities. They identified correlations between factors such as education, healthcare, and economic growth. The project also highlighted the impact of globalization on various countries and regions. By applying data science techniques, students developed critical thinking and problem-solving skills while gaining a global perspective, more specifically, the following:

- **Poverty and inequality:** Students analyzed poverty rates, income distribution, and social safety nets across countries.
- **Health disparities:** By examining factors like life expectancy, infant mortality, and healthcare spending, students identified global health challenges.
- **Economic growth and development:** Students explored the relationship between GDP, debt levels, and other economic indicators.
- **Education and human development:** By analyzing literacy rates, school enrollment, and educational attainment, students assessed the impact of education on societal progress.

Educational Impact

The "Global Socioeconomic Analysis" project significantly enhanced students' learning experience. By working with real-world data and addressing complex global issues, students developed a strong foundation in data science and gained a deeper appreciation for the field's potential to drive positive change:

- **Skill development:** Students acquired practical skills in data wrangling, visualization, modeling, and inference.
- **Critical thinking:** The project fostered critical thinking and problem-solving abilities as students analyzed complex data and drew informed conclusions.
- **Global awareness:** By exploring global disparities and trends, students developed a broader understanding of the interconnected world.
- **Interdisciplinary learning:** The project integrated data science with social sciences, providing students with a holistic perspective on global challenges.

Conclusions

The integration of globalization into data science education is essential for preparing students to address complex global challenges. The "Global Socioeconomic Analysis" project demonstrates the value of using real-world data to explore global issues and develop data science skills. By providing students with hands-on experience and opportunities for critical thinking, this project equips them to become informed and responsible global citizens.

To further enhance the impact of such projects, it is crucial to incorporate diverse perspectives, promote collaboration, and encourage students to communicate their findings effectively. By emphasizing the importance of data ethics and responsible data use, we can empower future data scientists to contribute to a more equitable and sustainable world.

Bibliography

Orit Hazzan & Koby Mike (2023), The Birth of a New Discipline: Data Science Education, in *Guide to Teaching Data Science*, pp. 59-72, Springer.

M Shad (2024), Data Science for Social Impact: Solving Real-world Problems, https://medium.com/@smr993875/data-science-for-social-impact-solving-real-world-problems-2186bf9f6b0d

Juraj Miština (2007), Globalization and its Impacts on Engineering Education, in Editors, Steen Hyldgaard Christensen, Martin Meganck, Bernard Delahousse, *Philosophy in Engineering*, pp. 381-408, Publisher: Academica, Denmark

Abhinav Rawat (2024), Data Science for Social Good: Solving Real-World Problems, https://www.interviewkickstart.com/blogs/articles/data-science-social-good-real-world-problems

Alfred Z. Spector, Peter Norvig, Chris Wiggins, Jeannette M. Wing (2023), *Data Science in Context: Foundations, Challenges, Opportunities,* https://datascienceincontext.com/wp-content/uploads/2022/10/Data-Science-in-Context-V.M1.pdf

Joseph Thomas (2023), Data Science for Social Impact: Changing the World, One Algorithm at a Time, https://www.linkedin.com/pulse/data-science-social-impact-changing-world-one-time-joseph-v-thomas-hxdhc/

Tiffany Timbers, Trevor Campbell, Melissa Lee, Joel Ostblom, Lindsey Heagy (2025), *Data Science: A First Introduction*, Routledge.