Data science is a relatively new field of study that is growing at an astronomical rate. Academic institutions are establishing data science programs, industries are hiring more and more data science graduates, and it is starting to integrate heavily with other professions. In the context of human development, data science provides a powerful avenue to improve understanding and inform important decisions to elevate intrinsic human freedoms. This essay will outline how data science can be used to challenge human development problems, areas where it has influenced understanding, and acknowledge the shortfalls of current and future approaches.

In West’s and Georgescu-Roegen’s works, socioeconomic systems undergo evolutionary processes that are analogous to natural phenomena. Furthermore, these systems are complex, meaning the constituents of the system only exhibit the observed behavior of the system when interacting as a whole. As humans, our ability to elucidate individual correlations from these many interactions is limited; this can be mitigated by data science techniques. Algorithms created as a result of knowledge in data science can more easily uncover observations that would have taken human minds a much longer time to deduce. This leads to the use of data science as a way to augment our scientific method (Kitchen). Given an initial data set, intelligently designed algorithms could potentially improve initial hypotheses. This would hopefully lead to more rapid discoveries and reduce resources spent on “false leads”. In regard to development, data science also has become massively interdisciplinary: the data produced from research can be sectioned into pieces which can be respectively analyzed by different experts. The combination of focused hypotheses and a collaborative environment could bolster efforts to understand development issues and further solutions.

Efforts in development research have already been influenced by data science methods. Researchers have been using mobile phones extensively to investigate behaviors that are consistent with different economic groups. Frequency of international phone calls and number of social media friends often indicates a wealthier status. Credit scores deduced from mobile phones can identify those that don’t have access to financial institutions (implying poverty). Additionally, mobile phones have also been used to provide microloans to this demographic. However, going back to our point of complexity, many of these simple solutions can have unforeseen consequences due to external factors that maintain the complex system. For example, microloans can make borrowers worse off due to unknown loan terms, faults in human integrity results in taking dishonest advantage of support systems, and a failure to consider temporal changes. Most importantly, the most vulnerable demographic of a population can be marginalized if data collection methods are biased.

Indeed, these problems hint at a bigger picture issue: the common view of data science as a “Janus Key” (from Roman myth). Many researchers view data science as a solution to all potential problems and proceed to use data science techniques with no knowledge of their consequences. True data science arguably requires a strong understanding of statistics and perhaps linear algebra. Most researchers new to data science do not have a formal education in data science and the associated mathematics. From this, it is easy to make false judgements of data and publish false results. From the previous paragraph, researchers failed to realize that their sample did not represent the population and was biased towards people who can afford mobile phones, effectively ignoring the poorest and most at-risk population.

A final obstacle that must be overcome is the radical idea to abandon theory in favor of algorithms. Theory is the fundamental basis upon which research is supported. Without it, research would be incredibly inefficient since researchers would not truly know what the data and their analysis mean. To elaborate, algorithms themselves are built upon theory testing and observations. Blind trust in algorithms would be the end-stage consequence of the point made in the previous paragraph - a lack of understanding will only harm more people.

In conclusion, data science is a promising new field that can potentially solve many issues present in human development. It has allowed researchers to accelerate their understanding of developing countries through rigorous scientific evaluation augmented by data science and through the interdisciplinary efforts of many. However, due to its status as a nascent field, the risk of false assumption is still prevalent. Researchers make fatal mathematical mistakes and treat statistical algorithms as black-boxes not worth understanding. It warrants the furtherance of formal academic programs to educate the next generation of data scientists to tackle these development problems. With these steps, I am confident data science will improve human existence.