towardsdatascience.com

Benefits and Ethical Challenges in Data Science — COMPAS and Smart Meters

Maximilian Hofer

22–28 minutes

Long-form essay.

Getting the balance right: weighing benefits against ethical challenges in Data Science.

1. Introduction

Recent advances in processing large amounts of data have generated a plethora of new opportunities to improve individual lives and the welfare of our societies. For example, smart meters monitoring domestic electricity consumption can help save electricity when not at home or suggest the cheapest electricity supplier depending on the consumption pattern.

However, imagine the electricity supplier's website states that if you provide real-time as opposed to weekly electricity consumption data, your yearly electricity costs would decline by 20 percent. Should electricity suppliers be allowed to pressure consumers into trading privacy for money?

1.1. Outline

This essay describes two applications of Data Science, namely the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) software system used in US courts and domestic smart meters, to contrast their benefits with ethical challenges.

1.2. Definitions of key terms

Defining Data Science has two components. First, vast amounts of data (terabytes, petabytes) are characterised by the three V's of Big Data: volume, velocity and variety (Laney, 2001). Second, Data Science stands for statistical models implemented in the form of software that can detect patterns in such data.

Ethical challenges arise when opinions on what is considered right and wrong diverge. For example, should an algorithm have the power to decide whether a defendant is released on bail or not? As statistical models are built on top of data, applications like COMPAS would require analysing how the data is generated in the first place. Ethical challenges might arise when police officers on patrol are controlling citizens in a biased way, but this goes beyond the scope of the essay.

To *improve individual lives and the welfare of our society* requires an application to create value in the lives of one or more individuals without negative externalities and within legal boundaries. In other words, someone *benefits* from using this technology.

Finally, a *smart meter* is an electronic device that measures domestic electricity consumption and communicates with other electronic devices as part of a connected network (think IoT).

2. Benefits and ethical challenges

COMPAS and smart meters make use of large amounts of data, provide clear and distinct benefits, raise compelling ethical challenges, are discussed by numerous scholars and appeared to have the highest present-day and future impact on society.

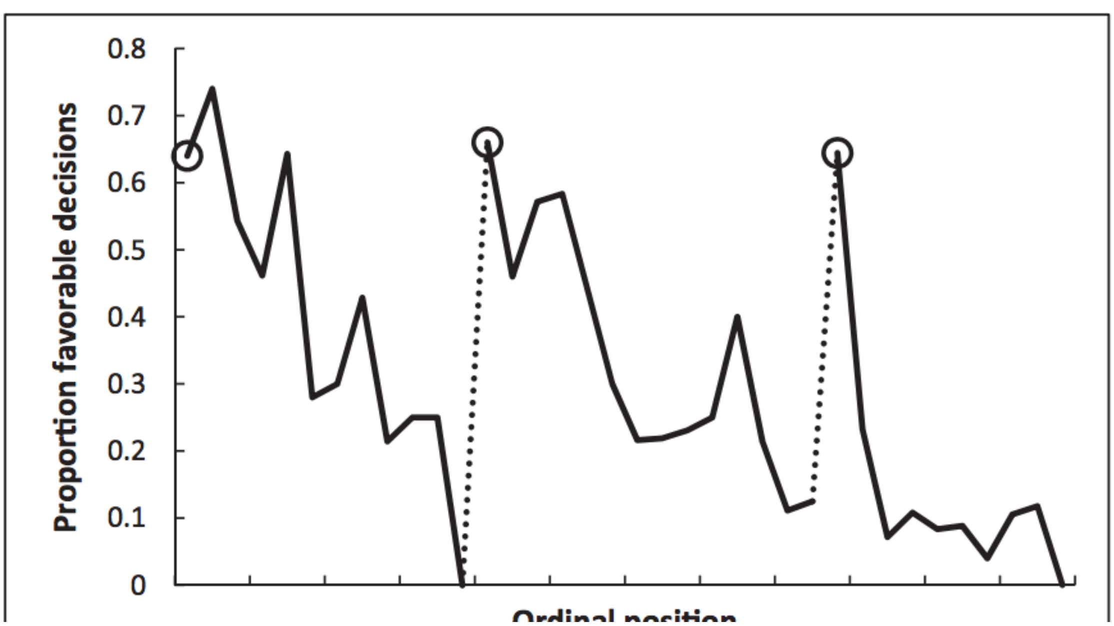
2.1. What is COMPAS?

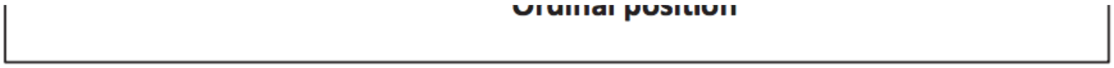
Judges across the United States are increasingly relying on algorithms to assess whether defendants awaiting trial are likely to re-offend. One such algorithmic system, developed by Northpointe

Inc., is called COMPAS and is used in Wisconsin, California, Colorado and Florida among others. Each defendant provides over 100 independent input variables during trial questioning, including age, sex and criminal history (Angwin, n.d.). Depending on how similar the defendant is to a particular group, COMPAS outputs risk scores that range from 0 to 10. Scores from 5 to 7 represent medium, scores from 8 to 10 high likelihood to re-offend. Judges consider scores as one of many factors in the decision-making process.

Benefits

Research papers such as Swets et al. (2000), have found data science methods to be superior to human judgement in certain settings. COMPAS complements human judgement, promising an objective and therefore fair assessment of a defendant based on all the case evidence available. It helps prevent outcomes, such as the one described by Danziger et al. (2011), who found that the likelihood of a favourable court rulings shoots up just after judges had meal breaks (see graph below).





Proportion of rulings in favour of the prisoners. Circled points represent the first decision after meal breaks (Danziger et al., 2011).

Furthermore, COMPAS shines in a criminal justice system cluttered with bureaucratic and administrative challenges by bringing ease of administration, brevity and efficiency to the decision-making process.

Ethical challenges

I identified three main ethical challenges: unfair discrimination, reinforcing human biases and lack of transparency.

Unfair discrimination.

What is a *fair* defendant evaluation and how can data scientists develop a corresponding algorithm that works for any defendant?

Northpointe’s definition of a fair algorithm states that the proportion of defendants who re-offend in each risk category is approximately the same regardless of race. In other words, a risk score of e.g. 7 predicts an equal likelihood of re-offending for black and white defendants. Any alternative definition would, by definition, discriminate against white defendants by artificially boosting their risk levels. Researchers from Stanford and Berkeley have validated that a single risk score represents an approximately equal risk of actual recidivism based on roughly 5,000 samples from Broward County, Florida (Corbett-Davies et al., 2016).

However, ProPublica, an investigative news organisation, identified unfair discrimination against black defendants by analysing over

10,000 defendant samples from the Broward County Sheriff's Office in Florida. They compared recidivism risk scores issued by COMPAS and actual recidivism rates two years after scores were issued and found that the algorithm classified black defendants as more than twice as likely to be considered medium to high risk than its white counterparts, even though neither black nor white defendants went on to commit a crime. Although black defendants did not re-offend, they were subject to harsher treatment by the court system (Larson et al., 2016).

Discriminatory treatment is not ethically problematic in itself; the effects of the treatment, i.e. being sent to jail or released on bail depending on race, are the ethical problems of this application (Schermer, 2011).

Can an algorithm mitigate ProPublica's legitimate concern while satisfying Northpointe's definition of fairness?

Reinforcing human biases.

Both aspects of Data Science defined earlier (vast amounts of data and statistical models) lead to an inherent limitation for making decisions based on patterns in past data. COMPAS' developers presumably fed the system with recidivism data to let the model identify variables correlating with a likelihood to re-offend — a standard way to train classification algorithms. If the data were perfectly unbiased there would be no problems, but all COMPAS can do is parrot back to us our own biases. These include severe, racist biases of judges that see black defendants as "more likely to kill again than whites" (Epps, 2017) or a Mexican defendant to have committed sexual assault "because he's Mexican and

Mexican men take whatever they want” (Supreme Court of Colorado, 2017).

Abdicating the responsibility for sentencing to a computer allows judges to make decisions based not only on relevant factors such as the seriousness of the offence but also on ethically problematic factors like inferred race and gender. Assume COMPAS predicts a defendant to be high risk. However, a judge disagrees and releases the defendant, who later re-offends. Why did the judge not comply with the software’s recommendation? This outcome is a lot easier to spot than when COMPAS suggests low risk, but the judge sentences the defendant and (probably) prevents recidivism.

Lack of transparency.

Hiding COMPAS’ inner workings from the public prohibits the understanding and discussion around mitigating learned biases. However, Northpointe understandably argues that publishing its proprietary algorithms would lead to a competitive disadvantage “because it’s certainly a core piece of our business” (Liptak, 2017) as an executive is quoted in the New York Times. This dilemma comes as no surprise when government agencies authorise private companies to apply Data Science to sensitive governmental data. There are two key areas requiring transparency.

First, transparency with regards to statistical modelling requires Northpointe to disclose the step-by-step process leading to risk scores, the underlying model and its parameters. Depending on the model(s) used, this task can be very difficult as the exact workings of models like neural networks are still unclear. Should lawmakers only allow algorithms that are fully understood to make such wide-reaching recommendations?

Second, it remains unclear what data is used. While race is not an input variable, other variables like gender are as made evident in the proceedings of Eric L. Loomis' case (Supreme Court of Wisconsin, 2016). It would be ethically very problematic if the dataset used for training could infer race based on other variables. With over 100 input variables, chances are that some subset of variables can accurately predict race. In that case, the statistical model cannot distinguish between the predictive power of a single variable and that set of variables. Northpointe's Chief Scientist Tim Brennan even admitted that "it is difficult to construct a score that doesn't include items that can be correlated with race" (Angwin et al., 2016). Sharing the correlation between variables does not conflict with trade secrets.

Should defendants have a claim right to know how COMPAS computes its score? Does COMPAS violate a defendant's right to an individualised sentence and right to be sentenced based on accurate information? ("State v. Loomis," 2017)

2.2. What are smart meters?

Recent advances in technology have triggered a shift towards more distributed power generation from many small-scale sources rather than few large-scale power plants. This requires at least a partial redesign of the grid, including smart meters, to guarantee secure operations. More specifically, system operators can gain network transparency and consumers can visualise and optimise their electricity consumption. Smart meters have numerous benefits but do not come without risks (Finster and Baumgart, 2015).

Benefits

Benefits for electricity suppliers can be split into network operations and more flexible billing (Jawurek et al., 2012).

Smart meters enable companies better grid management, including better projection of future network capacity requirements, preventative maintenance of network infrastructure and fast power outage detection (Depuru et al., 2011b).

With regards to billing opportunities, companies can leverage data to develop new tariff models, both for electricity consumption and generation (feed-in). This enables new markets with more choice for consumers. In fact, advanced metering infrastructure is a key enabler of a decentralised network with fluctuating electricity feed-in (Römer et al., 2012). Furthermore, smart meters can help detect fraud, reducing costs for both electricity suppliers and consumers. A study estimates that electricity suppliers worldwide lose approximately $25 billion a year due to electricity theft, which also causes higher carbon dioxide emissions due to higher electricity generation and missing funds to recover carbon dioxide. Even partial mitigation could lead to reduced costs for electricity companies and consumers and reduced carbon dioxide emissions, increasing the welfare of societies (Depuru et al., 2011a).

In addition to the above, consumers benefit from higher transparency into their electricity consumption, added flexibility to repurpose metering infrastructure for IoT hubs powering other smart services and new business models that e.g. allow charging electric vehicles away from home (Veale, 2017).

Ethical challenges

Ethical challenges can be categorised into three main areas: privacy, lack of transparency and consent and power.

Privacy.

At first glance, one might not categorise electricity consumption as particularly sensitive data. However, granular meter readings can be used to determine whether a person is at home or not, which appliances are used at what time (Molina-Markham et al., 2010), whether you leave appliances on for longer than required and even features of buildings (Beckel et al., 2014). Hence, intimate details of a user's daily life could be exposed and used in ways that invade individual privacy (Quinn, 2009).

Lack of transparency.

Opportunities for new tariffs, for example, raise a problem of transparency. Say my neighbour's and my tariffs vary even though we have the same flat size and data parameters. Is the difference due to me using more electricity on the weekends than she is? What data define tariff models?

Consent and power.

New tariff models would likely connect data granularity with tariff prices. In other words, the more granular the data, the cheaper the tariff. Hence the consumer can trade privacy for electricity costs, potentially pushing low-income household to consent to greater data sharing (Veale, 2017).

However, consumers need to understand and be able to access the data in the first place before giving consent to sharing. Nissenbaum (2011) has coined this issue the *transparency paradox*: low-level data is difficult to understand, and summary statistics hide crucial details — we cannot achieve one without giving up on the other.

3. Critical analysis

Instead of providing specific recommendations to overcome the challenges discussed, I would like to critically analyse the challenges discussed in part 2 with the following methods: statistical analysis, trade-offs rather than perfection, the ethics of data (individual and group privacy) and the ethics of practices.

3.1. COMPAS

Statistical analysis

Returning to the debate involving Northpointe and ProPublica, we have established two premises:

- The proportion of defendants who re-offend in each risk category should be approximately the same regardless of race (Northpointe's definition of fairness).
- Black defendants who do not re-offend are considered riskier than their white counterparts (ProPublica's criticism).

The missing piece of information is that the overall recidivism rate of black defendants is higher than the one of white defendants (52 percent vs. 39 percent) (Corbett-Davies et al., 2016). This implies that in order to satisfy Northpointe's definition of fairness, which we said was problematic to change, it is mathematically likely that for defendants who end up not re-offending, black defendants are assigned higher risk levels — the model represents the skew in the dataset.

This does not mean we should disregard ProPublica's point, but rather have an open discussion with stakeholders including lawmakers, judges and citizens about what algorithms should prioritise and what alternative policies could be introduced. New tracking technologies could render bail requirements useless so that no one is jailed unnecessarily, for example.

Trade-offs rather than perfection.

COMPAS does not make perfect decisions, but we cannot expect it to have zero bias and zero side-effects. Perfection is a direction to aim for, but when we pragmatically compare a court system with COMPAS to one without, the former seems to have a better ratio of benefits to challenges (Baase, 2012). We should still seek improvements, for example ending the partnership with Northpointe to develop a similar software tool inside the US government and prevent the dilemma of trade-secrets. The government could publicise the algorithmic details to ensure transparency.

3.2. Smart meters

Ethics of data.

The most pressing individual privacy challenges are reidentification and real-time monitoring. These can be mitigated by e.g. storing all data on-device until transmitting aggregate consumption data a day after consumption has occurred. This approach reduces informational value but increases privacy while still permitting the benefits discussed earlier. Electricity suppliers seek highly granular data, but no matter the consumer's income level, laws should ensure that data granularity cannot cross the line where personal dignity yields to the demands of electricity suppliers (Warren and Brandeis, 1890).

The European Commission (2014) has announced in 2014 that 16 out of 28 member-states will have wide-scale (80% or more) coverage of electricity smart meters by 2020. Critical topics like data granularity, security mechanisms and contingency planning have not yet been addressed.

Metering data can categorise people into groups according to particular consumption patterns (e.g. weekdays vs weekends), for example. Protecting individual privacy does not necessarily imply protecting group privacy. Groups of people can form attractive targets for criminals, e.g. citizens using less electricity on the weekends, suggesting they are not at home. Floridi has argued that the current policy frameworks (e.g. the current European legislation) are too anthropocentric (emphasising the natural person) and atomistic (only considering a single individual) to ensure group privacy. Laws protect individual privacy and the privacy of society, but not of groups within society. According Floridi (2015), such new laws would need to reconcile two moral duties: improving human welfare and fostering human right to privacy. Sometimes the only way to protect an individual is to protect the group, just like a fisherman with a shoal of sardines — he is trying to catch the shoal, not the individual sardine (Floridi, 2014).

Ethics of practices.

The traditional ethical analysis might propose a deontological code of conduct about what is ethical and what is not. However, we are in a situation of applied ethics, in which a consequentialist approach is more appropriate. This is because new technologies pose new problems, for which old methods have limited value. Policies can lead to unforeseen consequences, requiring a flexible policy framework that allows quick and frequent adaptation. One example of such a framework is the UK Minister for Cabinet Office (2016) *Data Science Ethical Framework* aimed at making innovation easier by establishing six key principles contrasting the public benefits against project risks. It is a simple and quick-to-comprehend document.

4. Conclusion

To conclude, this essay provides a critical analysing of the problem and the debate surrounding COMPAS and smart meters as examples of applying Data Science. The challenges identified include unfair discrimination, reinforcing human biases (COMPAS), privacy, consent and power (smart meters) and lack of transparency (both). The critical analysis used methods such as statistical analysis, trade-offs rather than perfection (COMPAS), individual and group privacy and the ethics of practices (smart meters).

Let us return to the question raised at the beginning: Should electricity suppliers be allowed to pressure consumers into trading privacy for money? *No*. Considering the methods discussed, policymakers need to grant a minimum level of privacy for consumers regardless of financial status. Finding the appropriate level needs an open public discussion (cf. COMPAS) and flexible policy frameworks (cf. smart meters).

What other applications come to your mind?

5. Bibliography

Angwin, J., n.d. Sample-COMPAS-Risk-Assessment-COMPAS-"CORE". ProPublica.

Angwin, J., Larson, J., Mattu, S., Kirchner, L., 2016. Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica.

Baase, S., 2012. A Gift of Fire: Social, Legal, and Ethical Issues for Computing Technology, 4 edition. ed. Pearson, Upper Saddle River, NJ.

Beckel, C., Sadamori, L., Staake, T., Santini, S., 2014. Revealing household characteristics from smart meter data. Energy 78, 397–410. https://doi.org/10.1016/j.energy.2014.10.025

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., 2016. A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. Wash. Post.

Danziger, S., Levav, J., Avnaim-Pesso, L., 2011. Extraneous factors in judicial decisions. Proc. Natl. Acad. Sci. 108, 6889–6892.

https://doi.org/10.1073/pnas.1018033108

Depuru, S.S.S.R., Wang, L., Devabhaktuni, V., 2011a. Electricity theft — Overview, issues, prevention and a smart meter based approach to control theft. Energy Policy, Special Section on Offshore wind power planning, economics and environment 39, 1007–1015. https://doi.org/10.1016/j.enpol.2010.11.037

Depuru, S.S.S.R., Wang, L., Devabhaktuni, V., Gudi, N., 2011b. Smart meters for power grid — Challenges, issues, advantages and status, in: 2011 IEEE/PES Power Systems Conference and Exposition. Presented at the 2011 IEEE/PES Power Systems Conference and Exposition, pp. 1–7. https://doi.org/10.1109/PSCE.2011.5772451

Epps, G., 2017. The Supreme Court Confronts Racism in the Jury Room. The Atlantic.

European Commission, 2014. Benchmarking smart metering deployment in the EU-27 with a focus on electricity.

Finster, S., Baumgart, I., 2015. Privacy-Aware Smart Metering: A Survey. IEEE Commun. Surv. Tutor. 17, 1088–1101. https://doi.org/10.1109/COMST.2015.2425958

Floridi, L., 2015. The Ethics of Information, Reprint edition. ed. Oxford University Press, Oxford.

Floridi, L., 2014. Open Data, Data Protection, and Group Privacy. Philos. Technol. 27, 1–3. https://doi.org/10.1007/s13347-014-0157-8

Jawurek, M., Kerschbaum, F., Danezis, G., 2012. SoK: Privacy Technologies for Smart Grids — A Survey of Options.

Laney, D., 2001. 3D Data Management: Controlling Data Volume, Velocity, and Variety.

Larson, J., Mattu, S., Kirchner, L., Angwin, J., 2016. How We Analyzed the COMPAS Recidivism Algorithm. ProPublica.

Liptak, A., 2017. Sent to Prison by a Software Program's Secret Algorithms. N. Y. Times.

Minister for Cabinet Office, 2016. Data Science Ethical Framework — GOV.UK.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L., 2016. The ethics of algorithms: Mapping the debate. Big Data Soc. 3, 2053951716679679. https://doi.org/10.1177/2053951716679679

Molina-Markham, A., Shenoy, P., Fu, K., Cecchet, E., Irwin, D., 2010. Private Memoirs of a Smart Meter, in: Proceedings of the 2Nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, BuildSys '10. ACM, New York, NY, USA, pp. 61–66. https://doi.org/10.1145/1878431.1878446

Nissenbaum, H., 2011. A Contextual Approach to Privacy Online. Daedalus 140, 32–48. https://doi.org/10.1162/DAED_a_00113

Quinn, E.L., 2009. Privacy and the New Energy Infrastructure (SSRN Scholarly Paper No. ID 1370731). Social Science Research Network, Rochester, NY.

Römer, B., Reichhart, P., Kranz, J., Picot, A., 2012. The role of smart metering and decentralized electricity storage for smart grids: The importance of positive externalities. Energy Policy, Special Section: Past and Prospective Energy Transitions — Insights from History 50, 486–495. https://doi.org/10.1016

/j.enpol.2012.07.047

Schermer, B.W., 2011. The limits of privacy in automated profiling and data mining. Comput. Law Secur. Rev. 27, 45–52. https://doi.org/10.1016/j.clsr.2010.11.009

State v. Loomis, 2017. . Harv. Law Rev.

Supreme Court of Colorado, 2017. Pena-Rodriguez v. Colorado (syllabus).

Supreme Court of Wisconsin, 2016. Eric L. Loomis v. Wisconsin.

Swets, J.A., Dawes, R.M., Monahan, J., 2000. Psychological Science Can Improve Diagnostic Decisions. Psychol. Sci. Public Interest 1, 1–26. https://doi.org/10.1111/1529-1006.001

Veale, M., 2017. Data management and use: case studies of technologies and governance — Produced for the British Academy and the Royal Society.

Warren, S.D., Brandeis, L.D., 1890. The Right to Privacy. Harv. Law Rev. 4, 193–220. https://doi.org/10.2307/1321160