The Sticky Wicket:
Unravelling the Hidden Politics in Cricket's Data Play

"Numbers have life; they're not just symbols on paper." - Shakuntala Devi

Yan Lin

Figure 1: Charlie Crowhurst/Getty Images for ECB

1. Introduction

Cricket, once the pastime of British aristocrats, has been colonised by an unseen player: data. As algorithms parse every swing, spin, and stride, we must contend with the imperial implications of data science that now governs this quintessentially English game.

1.1 Explanation

Critical perspectives in data science disclose the unseen forces at work in a field that is typically perceived as objective and neutral. They expose the power dynamics, biases and ideologies embedded in every stage of data work, from production and processing to use, allowing us to peel back the layers and go beyond numbers and algorithms to uncover the social, political, and ethical dimensions of data science (Didier et al., 2019). It compels us to recognise and investigate an often-overlooked reality: **data is not a neutral entity**.

Instead, it is embedded and influenced by human decisions, biases, and power structures at every stage of its lifecycle (Boyd & Crawford, 2012). Using the ethnographic study of the cricket academy as an example, these perspectives shed light on how the context within which data is collected and interpreted can shape the data itself (Bowles, 2014). Critical perspective not only informs the accuracy and reliability of data-driven predictions but also has ethical implications for the stakeholders involved (Eubanks, 2018).

1.2 Framework

This essay explores the political nature of data science, followed by a thorough literature review (Section 1.3) on the politics of data production, processing, and use, revealing concealed power dynamics and biases. The following (Section 2) describes my dissertation, which includes five machine learning models for predicting The Hundred Club cricket match outcomes. It then revisits the political implications and ethnographic perspectives in the context of my project (Section 3) during the politics of data production, processing, and use, investigating biases and power dynamics in the project's data and their broader social impacts, focusing on the under-analyzed Britishness and colonialism of the cricket data.

1.3 Literature Review

1.3.1 Data Production

The politics of data production reach into the realm of data classification, a process deeply influenced by socio-cultural contexts and power structures. This does not just mirror reality but shapes it, with census categories actively creating and reinforcing racial and ethnic divides (Bowker & Star, 2000). This shows how data, often considered objective, is shaped by societal influences, potentially leading to biases. Furthermore, categorising individuals and groups can have long-term impacts, leading to marginalisation, privileging certain knowledge, and reinforcing social categories (Hacking, 2015). As with categories of performance indicators in cricket (such as player statistics or team rankings), certain types of play or players may be favoured and rewarded more.

"Clean data", defined as accurate, comprehensive, and devoid of errors or biases, becomes an unrealistic ideal. Because the data is always "cooked" or manipulated in some way (Biruk, 2018). Data collection is frequently a messy process involving negotiations among researchers, participants, and the environment, as same as cricket match data. In addition, the idea of "raw data" is inherently contradictory (Gitelman, 2013). All data is contextually produced, thus revealing the politics of data production tied to power, control, and influence. This notion challenges the assumed neutrality of data.

However, "data vacuuming" refers to the practice of outsourcing the collection, analysis, and use of data to external organizations to meet the criteria of donor partners (Gimbel et al., 2018). This practice can lead to higher barriers for outsiders to access data within the cricket team. The consequence of such politics is a trend towards concentrating research on fewer, credible datasets (Koch et al., 2021). For instance, in cricket research, where the use of official datasets is prevalent. We can consider adopting the CARE principles of indigenous data governance (Carroll et al., 2020), which emphasise collective interest, control, accountability, and ethics, as a solution.

The intersection of data production and politics is brought into sharp focus in the work of Gimbel et al. (2018). "Uneven data infrastructures" draws attention to the global disparity in the distribution and accessibility of data infrastructures. In developed countries, data infrastructures tend to be more sophisticated, facilitating easier collection, analysis, and use of data. We can consider that the results and statistics of a hundred-ball cricket tournament that happened in England whilst another was held in Serbia are completely different.

Data colonialism is the decentralised extraction and control of citizens' data, with or without their explicit consent (Couldry & Mejia, 2019). This resembles the historical pattern of British colonialism, in which colonies were governed strictly to maintain control. Not only does Danielle (2019) emphasise this similarity, but she also highlights the limitations of data restriction laws. Cricket is a uniquely British sport, an integral element of British culture, and a manifestation of British colonialism in India, Australia, and globally (Malcolm, 2001). In the politics of data production, cricket is imbued with a colonial and inherently British identity.

1.3.2 Data Processing

In the politics of data processing, algorithms and models are crucial agents that can affect both societal structures and individual behaviour. Algorithms are cultural artefacts that embody and shape societal norms and political dynamics (Seaver, 2017). They have the power to undermine democratic processes and influence major societal decisions, emphasising the need for their design to be fair (Amoore, 2020).

Transparency is crucial for machine learning applications. Due to deliberate corporate secrecy, technical illiteracy, or the intrinsic opacity of machine learning algorithms and the necessary scale for their practical use, algorithmic opacity can impede societal understanding and promote misinformation (Burrell, 2016).

In addition, a prevalent misunderstanding is that computer models used in meteorological research are purely theoretical, devoid of observational grounding (Edwards, 2013). Another misconception is the assumption of infallibility and uncertainty of their predictions. Nevertheless, "think globally, act locally" (Edwards, 2013) asserts a close relationship between two very different scales: the macro, global environmental and economic system, on the one hand, and the microdomain of individual decisions and actions, on the other.

The proposition to incorporate solidarity as an ethical principle in AI signifies the need for a holistic socio-technical approach to its development (Luengo-Oroz, 2019). In fields like cricket prediction modelling, this necessitates a nuanced understanding of how algorithms and models, while invaluable for strategy development and performance evaluation, might inadvertently perpetuate biases and affect fairness (Buolamwini, 2018).

1.3.3 Data Use

Who uses data and why is crucial? Bowe et al. (2020) create counterplots and pandemic maps of marginalised groups to demonstrate data usage power. Williams (2020) emphasises ethical data use as a public good despite its private ownership.

Data feminism promotes rethinking binaries, pluralism, power, empowerment, and context, legitimising embodiment and affect, and making labour visible (D'lgnazio & Klein, 2016). This novel viewpoint on data utilisation highlights gender, social, and cultural variables and the importance of labour in data usage.

2. Dissertation Project Proposal

"What do they know of cricket who only cricket know? " -James, C. L. R.

2.1 Research Scope and Data Description

The title of my dissertation is "Can we predict the Outcome of cricket matches?" Given the extensive variety of cricket matches and the diversity in their rules, my project specifically concentrates on The Hundred competition, which was launched by the England and Wales Cricket Board (ECB) in 2021. In its two-year tenure, it has hosted 128 matches[1].

Each match consists of two innings and forty overs, with each team restricted to one hundred deliveries to amass their score (Nadal, 2019). The compact and rapid nature of this format implies that the data characteristics and game strategies will differ substantially from traditional cricket (Fletcher et al., 2023). The data is sourced from Cricsheet (n.d.)[2] and ESPNcricinfo (n.d.)[3]. However, it should be noted that the dataset is not The Hundred's official statistical record.

The dataset is split into two subsets: training data from the 2021 season and testing data from the 2022 season. Our goal is to predict the August 2023 match outcomes, including the winner and best player. This information includes but is not limited to, team wins and losses per match, detailed scores per inning, individual batting performances, toss decisions, etc. Each feature (variable) will undergo thorough scrutiny and processing. Given the limited training data, there is a potential risk of model "overfitting", meaning it may perform well on the training data but poorly on novel testing data (Ying, 2019).

2.2 Modelling Techniques and Implementation

Regarding modelling, I propose utilising a predictive classifier based on Naïve Bayes and Binomial Logistic Regression. Naïve Bayes is a classification algorithm based on the Bayesian framework that assumes all features are independent (Leung, 2007), even though this simplistic assumption is rarely met in practice. The Binomial Logistic Regression model, conversely, can handle interrelationships between features (Hilbe, 2009).

In addition to these models, I plan to introduce more complex algorithms, such as Decision Trees, which provide intuitive models. Given that our objective extends beyond accurate predictions to understanding why certain predictions are made, Decision Trees are a suitable choice (James et al., 2021). I also intend to incorporate Deep Neural Networks (DNN) and Support Vector Machines (SVM). DNNs represent some of the most powerful models in the data domain, capable of handling extremely complex patterns and relationships through layers of neurons and nonlinear transformations (Du, 2023). SVMs are algorithms that seek decision boundaries in high-dimensional spaces, boasting strong capabilities in handling high-dimensional features and avoiding overfitting (James et al., 2021).

---

[1] 32 group matches in 2021 and 2022 for men's cricket, 32 group matches in 2021 but only 24 in 2022 for women's cricket, plus the respective semi-finals and finals, totalling 128 matches.

[2] Cricsheet (n.d.) is a website built by cricket and statistics enthusiasts that offer free structured data on various types of cricket matches. Available at: https://cricsheet.org/downloads/.

[3] ESPNcricinfo (n.d.) is a sports news website exclusively dedicated to the game of cricket. Available at: https://www.espncricinfo.com/records/tournament/the-hundred-men-s-competition-2021-13880.

Models will be tested and evaluated scientifically. This includes using cross-validation to validate model robustness and stability (James et al., 2021) and a variety of performance indicators (accuracy, precision, recall, F1 score, AUROC, RMSE) (Passi & Pandey, 2018) to assess model performance (Du, 2023). Lastly, I'd like to develop a user-friendly online interface for prediction results or provide my source code. This interface would allow users to enter match information and receive a forecast based on the trained models.

3. Critical Perspective Analysis in My Project

3.1 Politics of Data Production

"It is not true that the English invented cricket as a way of making all other human endeavors look interesting and lively; that was merely an unintended side effect. " -Bill Bryson

3.1.1 Limited Representation

The missing data and subjectivity of the dataset result in limited representation. As stated previously, the data sources chosen were not the official statistical records of 'The Hundred' or ECB, but rather Cricsheet and ESPNcricinfo, which, despite being widely regarded as powerhouse sources, are nonetheless external, unofficial data sources. This process of data production and selection reflects a power dynamic and deterministic socio-cultural influences (Bowker & Star, 2000). To be more precise, the official matches consisted of 128 matches, but the dataset only recorded 124 (missing information on three matches and one weather-related suspension resulting in no winner[4]). We could add information on these four games to the library of the missing dataset (Onuoha, 2016). Besides the basic winner, team name, season, gender, venue, city, etc., the dataset contains variables such as toss winner, home advantage, toss decision, batting score, match number, and other English-specific variables. This implicitly reflects the British identity of cricket and the cultural background and preferences of the data collectors (Gitelman, 2013). In addition, crucial information, such as the weather and the athlete's physical condition, is omitted here. These variables are referred to as "cooked data" rather than "clean data"(Biruk, 2018). Cricket is not a biassed sport, but cricket data are not objective.

	Result	winner	team	opponent	gender	season	month	day	match_number	venue	...	toss_decision	choose_to_bat	choose_to_field	forced_to_bat	forced_to_field
0	loose	Oval Invincibles	Manchester Originals	Oval Invincibles	female	2021	7	21	1	Kennington Oval, London	...	bat	yes	no	no	no
1	win	Oval Invincibles	Oval Invincibles	Manchester Originals	female	2021	7	21	1	Kennington Oval, London	...	field	no	no	no	yes
2	win	Oval Invincibles	Oval Invincibles	Manchester Originals	male	2021	7	22	1	Kennington Oval, London	...	bat	no	no	yes	no
3	loose	Oval Invincibles	Manchester Originals	Oval Invincibles	male	2021	7	22	1	Kennington Oval, London	...	field	no	yes	no	no
4	loose	London Spirit	Birmingham Phoenix	London Spirit	female	2021	7	23	2	Edgbaston, Birmingham	...	bat	no	no	yes	no
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 2: A piece of the data variables I have read from this dataset.

[4] Details: Manchester Originals (Men) vs Southern Brave (Men). August 05, 2021. Available at: https://www.espncricinfo.com/series/the-hundred-men-s-competition-2021-1252040/manchester-originals-men-vs-southern-brave-men-19th-match-1252684/full-scorecard

Hidden discrimination leads to limited representation and inequality. Typically, the hundred requires 15 players per match, with a foreign player limit of four. In this year's format, however, there will be one less overseas position on each roster for both the men's and women's Hundred, with the overseas wildcard position no longer available (*The Cricketer*, 2023). When it comes to cricketers from other nations and regions, they are designated as "other" in the dataset, a labelling process that exacerbates the distribution of ethnicity and geography and reinforces the Eurocentric notion of justice (Barany, 2015). In 2018, racial discrimination at Yorkshire Cricket Club (*Wikipedia.*, n.d.) demonstrates that there are still entrenched racial issues in cricket. During the production of data, this racial discrimination may be perceptible and silent, generating implicit racism (Dimeo & Kay, 2004). This implies not only ethnic bias but also gender prejudice. In fact, the highest salary for a male player on the hundred cricket team is £125,000, whilst the maximum salary for a female player is only £31,250. The concept of global feminism emphasises the need to examine the gender pay disparity in a broader social and cultural context (Fullagar & Pavlidis, 2012), which sheds further light on the significance of examining the politics of data production.

3.1.2 Data Colonialism and Power Dynamics in Cricket

From an ethnographic perspective, data production in cricket is the recording and cataloguing of a cultural performance (Bale, 2002). Moreover, cricket's data production is determined by its British-centric norms and colonial history, a manifestation of the concept of data colonialism proposed by Couldry & Mejias (2018). Historically, Britain used cricket to "civilise" its colonies, such as India and Australia, imparting "British values" and leaving an indelible mark on the sport (Fletcher, 2011). This British orientation of cricket, including its metrics, structures, and norms, dictates data production politics. The Hundred competition, which is contested exclusively in seven British cities, exemplifies this by utilising a substantial amount of player and match data. Understandings of performance quality are established by the dominant metrics influenced by British standards. This can marginalise non-British viewpoints and methods, such as the "doosra" (Wikipedia., n.d.), perpetuating power inequalities and limiting representation (Bourdieu, 1977).

The politics of data production in cricket disclose power dynamics comparable to colonialism, both historical and digital. Organisations that collect, control, and analyse data can influence the narrative of a game. This may perpetuate data colonialism by imposing British norms and standards and marginalising other forms of cricketing expertise and knowledge. This critique highlights the importance of addressing the politics of data production to ensure a more equitable representation of diverse cricketing nations, styles, and abilities, thereby challenging the continuation of data colonialism within the sport (Couldry & Mejias, 2019).

3.2 Politics of Data Processing: Algorithmic Bias

It is essential to be aware of potential algorithmic biases when employing a variety of models (Naive Bayes, Binary Logistic Regression, and others) for cricket prediction. According to Buolamwini (2018), algorithms may unintentionally perpetuate bias and unfairness. In sports prediction, bias might occur if the models overly favour certain variables (such as, "home advantage" and "toss decision", with weights and interaction terms) that could unfairly benefit or disadvantage some players or teams.

Random Forests which have shown promising results in similar sports prediction (Passi & Pandey, 2018). Though their ability to manage high-dimensional data and avoid overfitting is commendable, their opacity could lead to "black box" issues. The lack of interpretability may cloak potential biases in the model's decision-making process (Burrell, 2016). SVM's reliance on the chosen kernel (James et al., 2021) and DNN's susceptibility to overfitting could influence the model's fairness (Lotfollahi et al., 2019).

3.3 Politics of Data Use: Preventing Misuse

Schools, cricket forums, and The Hundred officials have access to my final prediction model; therefore, data should be treated as a public good (Williams, 2020), ensuring that it is not exploited for personal or corporate gain, particularly in terms of betting or gambling in cricket matches.

Pasquale (2015) and Hao (2021)'s insights highlight the need to ensure that the data and AI in my prediction model are not misused or lead to unjust consequences, like potentially unfair betting practices. Regular audits and ethical guidelines could be one method for mitigating such risks.

4 Conclusion

In conclusion, this essay probes the fascinating world of cricket data, employing sophisticated algorithms to forecast the outcomes of The Hundred cricket matches. As we traverse the data journey from production to use, we encounter a pitch layered with societal, historical, and technological complexities.

A focus is placed on the 'Britishness' and colonial legacy embedded in cricket data, a seldom-explored aspect. Other obstacles such as underrepresentation due to incomplete datasets, various forms of veiled discrimination, and ethical quandaries surrounding the misuse of data are also discussed. This essay advocates for an enlightened and inclusive approach to employing data for insightful and equitable applications in sports.

Words: 2721 (25+1118+511+1067)

Reference

Didier Bigo, Isin, E.F. and Evelyn Sharon Ruppert (2019). Data politics : worlds, subjects, rights. Abingdon, Oxon ; New York, Ny: Routledge.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society, 15(5), 662-679.

Bowles, Harry (2014). “Days in the Dirt”: An ethnography on cricket and self. Cardiff Metropolitan University. Thesis. https://doi.org/10.25401/cardiffmet.20319828.v1

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Bowker, G.C. and Susan Leigh Star (2000). Sorting things out: classification and its consequences. Cambridge, Massachusetts: Mit Press.

Hacking, I. (2015). Biopower : Foucault and beyond. Chicago ; London: The University Of Chicago Press.

Cal (Crystal) Biruk (2018). Cooking Data. Duke University Press Books.

Gitelman, L. (2013). Raw data is an oxymoron. Erscheinungsort Nicht Ermittelbar: Verlag Nicht Ermittelbar.

Gimbel, S., Chilundo, B., Kenworthy, N., Inguane, C., Citrin, D., Chapman, R., Sherr, K. and Pfeiffer, J. (2018). Donor data vacuuming. Medicine Anthropology Theory, 5(2). doi:https://doi.org/10.17157/mat.5.2.537.

Koch, B., Denton, E., Hanna, A. and Foster, J.G. (2021). Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research. arXiv:2112.01716 [cs, stat]. [online] Available at: https://arxiv.org/abs/2112.01716.

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. Data Science Journal, 19(1). doi:https://doi.org/10.5334/dsj-2020-043.

Couldry, N. and Mejias, U.A. (2018). Data Colonialism: Rethinking Big Data’s Relation to the Contemporary Subject. Television & New Media, 20(4), pp.336–349.

Coleman, D. (2019). Digital Colonialism: The 21st Century Scramble for Africa through the Extraction and Control of User Data and the Limitations of Data Protection Laws. Michigan Journal of Race & Law, 24(24.2), p.417. doi:https://doi.org/10.36643/mjrl.24.2.digital.

Malcolm, Dominic. (2001). 'It's Not Cricket': Colonial Legacies and Contemporary Inequalities. Journal of Historical Sociology. 14. 253 - 275. 10.1111/1467-6443.00146.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. Big Data & Society, 4(2), pp.1–12. doi:https://doi.org/10.1177/2053951717738104.

Amoore, L. (2020). Cloud ethics: algorithms and the attributes of ourselves and others. Durham: Duke University Press.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), p.205395171562251. Doi: https://doi.org/10.1177/2053951715622512.

Edwards, P.N. (2013). A vast machine: computer models, climate data, and the politics of global warming. Cambridge, Mass.: Mit Press.

Luengo-Oroz, M. (2019). Solidarity should be a core ethical principle of AI. Nature Machine Intelligence, 1(11), pp.494–494. doi:https://doi.org/10.1038/s42256-019-0115-3.

Buolamwini, J. and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. pp.77–91.

Bowe, E., Simmons, E. and Mattern, S. (2020). Learning from lines: Critical COVID data visualizations and the quarantine quotidian. Big Data & Society, 7(2), p.205395172093923. doi:https://doi.org/10.1177/2053951720939236.

Williams, S. (2020). Data Action. MIT Press.

Klein, L.F. and D'Ignazio, C. (2016). Data feminism. Cambridge, Massachusetts: The Mit Press.

Nadal, S. (2019). The Hundred: Your guide to cricket's new quickfire competition. [online] Sky News. Available at: https://news.sky.com/story/the-hundred-your-guide-to-crickets-new-quickfire-competition-11840168.

Fletcher, T., Sturm, D. and Malcolm, D. (2023). A 'cannibalised' cricket event? Mediatisation, innovation and The Hundred. Leisure Studies, pp.1–16. doi:https://doi.org/10.1080/02614367.2023.2183980.

Ying, X. (2019). An Overview of Overfitting and its Solutions. Journal of Physics: Conference Series, 1168(2), p.022022. doi:https://doi.org/10.1088/1742-6596/1168/2/022022.

Leung, K. (2007). Naive Bayesian Classifier. Retrieved from: https://cse.engineering.nyu.edu/~mleung/FRE7851/f07/naiveBayesianClassifier.pdf

Hilbe, J.M. (2009). Chapter 4. Logistic regression models. In Logistic regression models (pp.63-72). London: Crc Press.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning: with Applications in R (2nd ed.). Springer. (pp. 197-223, 327-402)

Du, D.H. (2023). Machine Learning Part II. Durham University. Retrieved from: https://bookdown.org/hailiangdu/Lecture_notes/

Passi, K. and Pandey, N. (2018). Increased Prediction Accuracy in the Game of Cricket Using Machine Learning. International Journal of Data Mining & Knowledge Management Process, 8(2), pp.19–36. doi:https://doi.org/10.5121/ijdkp.2018.8203.

MIMI Onuoha. (2016). The Library of Missing Datasets — MIMI ỌNỤỌHA. [online] Available at: https://mimionuoha.com/the-library-of-missing-datasets.

The Cricketer (2023). ECB freeze Hundred salaries for 2023 in both men’s and women’s competitions | The Cricketer. [online] Available at: https://www.thecricketer.com/Topics/thehundred/ecb_freeze_hundred_salaries_2023_mens_womens_competitions.html#:~:text=Women.

Barany, M. J. (2015) “Ian Hacking, Why Is There Philosophy of Mathematics at All? Cambridge: Cambridge University Press, 2014. Pp. xv 290. ISBN 978-1-107-65815-8. £17.99 (paperback).,” The British Journal for the History of Science. Cambridge University Press, 48(4), pp. 686–687. doi: 10.1017/S0007087415000692.

Wikipedia. (2022). Yorkshire County Cricket Club racism scandal. [online] Available at: https://en.wikipedia.org/wiki/Yorkshire_County_Cricket_Club_racism_scandal.

Paul Dimeo & Joyce Kay (2004) Major sports events, image projection and the problems of 'semi-periphery': A case study of the 1996 South Asia cricket World Cup, Third World Quarterly, 25:7, 1263-1276, DOI: 10.1080/014365904200281267

Fullagar, S & Pavlidis, A. (2012) ‘It’s all about the journey’: women and cycling events,

International Journal of Event and Festival Management, Vol. 3 Iss: 2 pp. 149 – 170.

http://dx.doi.org/10.1108/17582951211229708

Bale, J. (2002). Sports geography. London: Routledge.Couldry, N., Ulises Ali Mejias and Stanford University Press (2019). The costs of connection: how data is colonizing human life and appropriating it for capitalism. Stanford: Stanford University Press.

Fletcher, T. (2011). The making of English cricket cultures: empire, globalization and (post) colonialism. Sport in Society, 14(1), pp.17–36. doi:https://doi.org/10.1080/17430437.2011.530006.

Wikipedia. (n.d.). Doosra. [online] Available at: https://en.wikipedia.org/wiki/Doosra

Bourdieu, P. (1977) Outline of a Theory of Practice. Translated by R. Nice. Cambridge: Cambridge University Press (Cambridge Studies in Social and Cultural Anthropology). doi: 10.1017/CBO9780511812507.

Lotfollahi, M., Jafari Siavoshani, M., Shirali Hossein Zade, R. and Saberian, M. (2019). Deep packet: a novel approach for encrypted traffic classification using deep learning. Soft Computing. doi:https://doi.org/10.1007/s00500-019-04030-2.

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Hao, K. (2019). AI Is Sending People to Jail—and Getting It Wrong. [online] MIT Technology Review. Available at: https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/.