Alternative Credit Scoring – A Double Edged Sword

—Shehnaz Ahmed*

Access to finance is a major challenge impeding small businesses and entrepreneurs. It is perceived that the traditional credit scoring framework fails to adequately capture 'thin file' borrowers who may have little or no credit history. However, technological innovation has enabled the use of big data to assess the creditworthiness of 'thin file' customers who lack detailed credit history on the basis of non-traditional data sets such as telecom data, social media data, buyers' patterns on e-commerce websites, etc. This has led to the emergence of lending models that rely on such alternative credit scoring mechanisms. However, the use of big data for algorithmic scoring also poses certain challenges which include concerns regarding discriminatory practices, data privacy, transparency, etc. With several studies recommending the use of digital lending coupled with alternative credit scoring for the MSME sector, these issues assume immense significance.

Keywords: *Lending; Data Protection; Credit Scoring; Machine Learning; Data Fiduciaries;*

I. Introduction

A well-functioning credit information system is an essential part of the financial infrastructure. It addresses asymmetry of information between borrowers and lenders in credit markets that leads to 'adverse selection' and moral hazard problems.[1] It can reduce the risk of a default and improve the allocation of credit to those who have been traditionally unserved or underserved by the formal credit system. One of the earliest disruptions in the credit information ecosystem was the introduction of credit scoring to assess the creditworthiness of prospective

* Senior Resident Fellow and Lead, Fintech at Vidhi Centre for Legal Policy, New Delhi.

[1] 'Disruptive Technologies in the Credit Information Sharing Industry: Developments and Implications' (*World Bank*, 2019) 1 <http://documents.worldbank.org/curated/en/587611557814694439/pdf/Disruptive-Technologies-in-the-Credit-Information-Sharing-Industry-Developments-and-Implications.pdf> accessed 2 January 2020.

loan applicants.[2] Traditionally, financial institutions relied on the judgment of lending officials for evaluating individual consumers for credit purposes. Based on the specific criteria (such as employment, income, age, etc.), these officials exercised wide discretion to assess the credit risk of a prospective consumer.[3] This left the lender relying on the business judgment of such officials to approve loan applicants. However, with digitisation, this manual system was replaced by automated credit scoring systems. These systems relied on algorithms that assessed the information provided by the applicant and other sources by using a set of weighted variables which computed the credit score of the applicants.[4]

In India, most lenders rely on credit scores (such as the CIBIL score) computed by Credit Information Companies ("**CICs**")[5] to assess the creditworthiness of prospective borrowers. A credit score in India is a 3-digit numeric summary of an individual's credit history that may be approximately around 300 to 900.[6] Most lenders prefer giving loans to borrowers with a credit score of 700 or above.[7] Typically, CICs compute credit scores using historical financial data such as defaults on prior credit transactions, payment behaviour and previous searches.[8] Different parameters such as length of credit history, number of defaults, recent repayment trends, amount and number of accounts overdue, presence of write-offs, the mix of credit, number of loan enquiries made, occupation and income etc., are applied towards determining a credit score.[9] Thereafter different weightage is assigned to each factor or parameter.[10] However, there is

---

2 ibid 2.

3 Kenneth G Gunter, 'Computerised Credit Scoring's Effect on the Leading Industry' (2000) 4 NC Banking Inst 443.

4 ibid.

5 'Applying for loan? Know your credit score first' *The Economic Times* (2 May 2013) <https://economictimes.indiatimes.com/wealth/borrow/applying-for-loan-know-your-credit-score-first/articleshow/19831152.cms> accessed 3 January 2020; 'How can you start building a credit score if you have never taken a loan' (*Money Control*, 13 November 2018) <https://www.moneycontrol.com/news/business/personal-finance/how-you-can-start-building-a-credit-score-if-you-have-never-taken-a-loan-3154861.html> accessed 3 January 2020.

6 'Understanding your CIBIL Score' *(CIBIL)* <https://www.cibil.com/resources/docs/CIBIL-Report-Understanding.pdf> accessed 2 January 2020; RBI, Data Format for Furnishing of Credit Information to Credit Information Companies and other Regulatory Measures, RBI/2014-15/128 (Notified on 15 July 2014).

7 'What is the Difference Between CIBIL Score and CIBIL Report?' (*ICICI Bank*) <https://www.icicibank.com/blogs/credit-card/what-is-the-difference-between-CIBIL-score-and-CIBIL-report.page?> accessed 2 January 2020.

8 Aditya Puri Committee, Report of the Committee to Recommend Data Format for Furnishing of Credit Information to Credit Information Companies, 67(January 2014).

9 ibid 68.

10 For instance, it is reported that a weightage of 35% of the total credit score is assigned to payment history, 30% to total credit, 15% to the length of credit and 10% is assigned to new credit and demographics such as age, location of residence, etc. Please *see*, 'Credit score calculation: The biggest secret revealed', *The Financial Express* (23 June 2017) <https://www.financialexpress.com/money/the-biggest-secret-revealed-heres-how-your-credit-score-is-calculated/732144/> accessed 2 January 2020.

no single or universal formula to compute a credit score, and different CICs in practice employ different methodologies. For instance, the CIBIL score prepared by TransUnion CIBIL Ltd. considers four significant factors *viz.* payment history, high credit utilisation, recently sanctioned loans and credit cards, and a mix between secured and unsecured loans.[11] CICs obtain such information from the financial and credit-related information submitted to them by lending institutions under the Credit Information Companies (Regulation) Act, 2005 ("**CIC Act**"), that is the primary legislation governing CICs in India.

As the conventional credit scoring models mostly rely on the past financial information of borrowers, it is perceived to be inadequate to capture borrowers with a thin credit file.[12] However, the development of technological innovations has paved the way for credit scoring systems to exploit a wide variety of non-traditional data, such as social media footprints, online behavioural data, telecommunications data, digital payments data, and calling patterns and contacts.[13] For instance, data about payments received or sent by mobile phone can be used to predict a person's income and her ability to repay.[14] Disruptive technologies have enabled these institutions to leverage big data and machine learning for drawing insights from online consumer behaviour and predicting her conduct regarding the transaction. However, the lack of transparency in the operation of such alternative credit scoring systems to assess creditworthiness raises concerns about such systems and their potential to produce arbitrary and discriminatory results.

Against this background, this article discusses the evolution of alternative credit scoring systems in India and the major concerns that these systems pose. It also studies the adequacy of the existing legal framework in India to address such concerns. In doing so, the article briefly attempts to provide a possible regulatory response that may be considered to leverage the potential of such alternative credit scoring systems and to contain their risks.

---

11 'Understanding your CIBIL Score' *(CIBIL)* <https://www.cibil.com/resources/docs/CIBIL-Report-Understanding.pdf> accessed 2 January 2020.

12 CIBIL score is derived based on the credit history of the applicant. *See* 'Score Simulator' (*CIBIL*) <https://myscore.cibil.com/CreditView/support.page?tab=faq&_ga=2.211853591.2027198589.1581765877-2002625815.1581765877> accessed 10 February 2020 ; *see also*, 'Credit Reporting Knowledge Guide 2019'(*World Bank*, 2019) <http://documents.worldbank.org/curated/en/262691559115855583/pdf/Credit-Reporting-Knowledge-Guide-2019.pdf> accessed 10 February 2020.

13 World Bank (n 1) 19; Eva Wolkowitz, Sarah Parker, 'Big Data, Big Potential: Harnessing Data Technology for the Underserved Market' (2015) 7, 27 <https://s3.amazonaws.com/cfsi-innovation-files/wp-content/uploads/2017/02/13062352/Big-Data-Big-Potential-Harnessing-Data-Technology-for-the-Underserved-Market.pdf> accessed 3 January 2020.

14 Tobias Baer, Tony Goland, and Robert Schiff, 'New credit-risk models for the unbanked' (April 2013) <https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Risk/Our%20Insights/New%20credit%20risk%20models%20for%20the%20unbanked/New%20credit%20risk%20models%20for%20the%20unbanked.ashx> accessed 2 January 2020.

II. The Rise of Alternative Credit Scoring Models

The advancements in technological innovations have equipped credit scorers with new-age tools to leverage non-traditional data or alternative data. The definition of such alternative data is likely to evolve with technological developments. However, it may be broadly defined to include data from non-traditional and non-financial sources such as messaging content, social media footprints, digital payments history, online browsing behaviour telecommunications data and locational data.[15] While traditional credit scoring models rely on structured data, alternative data may be both structured and unstructured.[16] Broadly, structured data is organised in a manner that makes it easily readable and searchable by algorithms.[17] It includes data about credit history, payments and assets.[18] On the other hand, unstructured data is not organised in a specific manner[19] and can include data relating to social media use, emails, text messages, mobile usage and locational data.[20] The advancements in machine learning enable exploitation of unstructured data by the new age alternative credit scorers to identify trends and patterns in consumer behaviour, including assessing the consumers' capacity and willingness to repay loans.[21] These credit assessment models also create opportunities for lenders to develop such financial products and services that are better mechanisms to address the need of the consumer.

Incumbent players (such as CICs) have exploited these technological disruptions, and it has also paved the way for new business models. For instance, reports indicate that Equifax Credit Information Services Private Limited, a registered CIC, has set up a separate analytic wing, which has partnered with fintech[22]

---

15 World Bank (n 1) 19-21; International Committee on Credit Reporting (ICCR), 'Use of Alternative Data to Enhance Credit Reporting to Enable Access to Digital Financial Services by Individuals and SMEs operating in the Informal Economy Guidance Note', ¶ 7, ¶ 17, ¶ 21, <https://www.gpfi.org/sites/gpfi/files/documents/Use_of_Alternative_Data_to_Enhance_Credit_Reporting_to_Enable_Access_to_Digital_Financial_Services_ICCR.pdf> accessed 2 January 2020.

16 ICCR (n 15) ¶ 23.

17 World Bank (n 1) 8; 'Financial Consumer Protection and New Forms of Data Processing Beyond Credit Reporting' (*World Bank*, November 2018) 3 <http://documents.worldbank.org/curated/en/677281542207403561/pdf/132035-WP-FCP-New-Forms-of-Data-Processing.pdf> accessed 2 January 2020.

18 ibid.

19 ICCR (n 15) ¶ 25.

20 ibid.

21 ibid.

22 The Financial Stability Board has defined the term 'fintech' as 'technology-enabled innovation in financial services that could result in new business models, applications, processes or products with an associated material effect on the provision of financial services.' Fintech firms and companies used here describe firms and companies whose business model focuses on these innovations. *See* 'FinTech and market structure in financial services: Market developments and potential financial stability implications' (*Financial Stability Board*, 14 February 2019)<https://www.fsb.org/wp-content/uploads/P140219.pdf> accessed 10 February 2020.

companies "to provide customers analysis for banks."[23] Similarly, the lending ecosystem in India that has traditionally been dominated by banks and non-banking financial companies ("**NBFCs**") is witnessing the emergence of new players and partnerships between incumbents and fintech companies. New and innovative business models in the digital lending space include peer to peer lending platforms, bank-fintech partnership models, NBFC-fintech partnership and online market places selling financial products.[24] Driven by technological innovations, these models primarily rely on consumer data to provide innovative services and products, including credit risk assessment models. While the alternative credit scoring ecosystem is still evolving in India, such models have already been operating in other jurisdictions for quite sometime. This part seeks to highlight how these new-age credit scorers exploit alternative data with specific examples of such-business models.

CreditVidya, a Hyderabad based company relies on alternative data, traditional credit reports (generated by CICs) and uses artificial intelligence-based algorithms to compute credit scores.[25] The 'CV Score' generated by CreditVidya takes into account 10,000 data points from applicants' SMSes (only commercial, not personal), such as utility bill payments, e-commerce transactions and phone-related location data.[26] This score is then relied on by lender partners (such as banks and NBFCs) for providing loans. The company's proprietary algorithm analyses unstructured data, and based on the risk appetite of the lender, either approves or rejects the application.[27] Similarly, Early Salary, an online lending platform, provides short financing solutions to individuals. The platform uses an

23 'Alternative data can gauge creditworthiness, if law permits', *The Economic Times* (6 April 2019), <https://economictimes.indiatimes.com/markets/stocks/news/alternative-data-can-gauge-creditworthiness-if-law-permits/articleshow/68748552.cms?from=mdr> accessed 2 January 2020.

24 Boston Consulting Group, 'Digital Lending A $1 Trillion Opportunity Over the Next 5 Years' (*Boston Consulting Group*, 2018) 18–20 <http://image-src.bcg.com/Images/BCG-Digital-Lending-Report_tcm9-197622.pdf> accessed 2 January 2020; 'Lending: from Paytm to Zerodha to Ola Money, fintech players find a new way to stay in the game', *The Economic Times* (4 September 2019) <https://prime.economictimes.indiatimes.com/news/70969179/fintech-and-bfsi/lending-from-paytm-to-zerodha-to-ola-money-fintech-players-find-a-new-way-to-stay-in-the-game> accessed 2 January 2020; Accion, 'Demystifying Digital Lending' (April 2018) <https://www.microfinancegateway.org/sites/default/files/publication_files/1123_digital_lending_r10_print_ready.pdf> accessed 2 January 2020; International Finance Corporation (IFC), 'Financing India's MSMEs', 7682 (November 2018), <https://www.intellecap.com/wp-content/uploads/2019/04/Financing-Indias-MSMEs-Estimation-of-Debt-Requireme-nt-of-MSMEs-in_India.pdf> accessed 2 January 2020.

25 The Software Development Kit of CreditVidya is integrated with the android applications, which enables the lenders to collect consent driven data. This unstructured data is then processed on CreditVidya's platform to allow the lender to assess the credit risk of the applicant, 'even when they have scanty or no bureau credit scores'. *See* 'Learn more about how we are raising the bar for Data Security & Privacy (What does CreditVidya's SDK do?)' (*CreditVidya*) <https://creditvidya.com/faqs> accessed 3 January 2020.

26 'CVScore Reimagining credit for the underserved' (*CreditVidya*) <https://creditvidya.com/static/pdf/CVScoreCaseStudy.pdf> accessed 2 January 2020.

27 ibid.

algorithm "that combines traditional credit scoring with new social and online scoring technology-linked risk assessment concepts."[28] The platform also collects SMS and browsing history, information from social networking platforms, such as Facebook, LinkedIn, and 'other similar platforms.'[29] Customers are required to do a social media login since the platform uses the internet presence of customers on "social networking sites as a reliable benchmark to assess creditworthiness."[30] CASHe, another online lending platform, provides short-term personal loans to 'young salaried professionals.'[31] The platform uses a proprietary predictive algorithm[32] called the 'Social Loan Quotient'[33] to create credit profiles of users using alternative data, such as mobile and social media footprint, education, monthly salary and career experience.[34] The company describes the Social Loan Quotient as a dynamic score that evolves with more interactions.[35] While the financial institutions provide the loans on their books, the technology companies offer necessary technical support to the partner lending institution to digitise the loan underwriting process.

Given that the alternative credit scoring ecosystem is still evolving in India, it may be worthwhile to study some models in other jurisdictions to understand the nature of this model better. One of the more prominent players in the alternative credit scoring ecosystem globally is ZestFinance. Inspired by the philosophy that "all data is credit data",[36] the company combines traditional credit

---

28 'Our mobile first strategy will help customers get credit in minutes' <https://www.earlysalary.com/about-us> (*EarlySalary*) accessed 2 January 2020.

29 'Privacy Policy' *(EarlySalary*, 6 July 2017) <https://www.earlysalary.com/privacy-policy> accessed 2 January 2020.

30 'Frequently Asked Questions, Application Process, Why do I have to do a social login?' (*EarlySalary*) <https://www.earlysalary.com/FAQ> accessed 2 January 2020.

31 'CASHe – Instant Personal Loan App' *(CASHe,*) <https://www.cashe.co.in/> accessed 3 January 2020.

32 An algorithm is a formally specified sequence of logical operations that provides step- by-step instructions for computers to act on some kind of input data and thus automate decision. *See* Solon Barocas et al, 'Data & Civil Rights: Technology Primer' (2014) <http://www.datacivilrights.org/pubs/2014-1030/Technology.pdf> accessed 10 February 2020. A proprietary predictive algorithm discussed above refers to an algorithm specifically designed by a company for predicting consumer behaviour (in the instant case probability of default in loan repayment) for lending decisions.

33 'Instant Personal Loans at your Fingertips' *(CASHe*, 3 August 2017) <https://www.cashe.co.in/personal-loans-at-your-fingertips/> accessed 3 January 2020.

34 'Social Loan Quotient' (*CASHe*) <https://www.cashe.co.in/about/> accessed 3 January 2020.

35 'Here are Answers to all the Questions that you were Shy to Ask Us, What is Social Loan Quotient' (*CASHe*) <cashe.co.in/faq/> accessed 2 January 2020.

36 'Just the Facts. Yes, All of Them', *The New York Times* (25 March 2012) <https://archive.nytimes.com/query.nytimes.com/gst/fullpage-9A0CE7DD153CF936A15750C0A9649D8B63.html> accessed 3 January 2020; 'Through an App, Darkly: How Companies Construct Our Financial Identity', *(Privacy International*, 30 January 2018) <https://privacyinternational.org/news-analysis/1099/through-app-darkly-how-companies-construct-our-financial-identity> accessed 3 January 2020.

information with alternative data based on online and offline activities of borrowers.[37] The company's proprietary algorithm analyses several thousand data points of an individual to arrive at a final score.[38] The exact variables used and the proprietary algorithm employed by ZestFinance to compute the credit score is a well-guarded secret. However, it has been reported that some of the factors that the company considers are whether an applicant typed her name in upper or lowercase, whether she has read the letter on the company's website, whether the applicant has a prepaid or postpaid phone connection, and how she spends her money *vis-a-vis* the expected patterns in the kind of city or town she lives in.[39] A review of a patent granted to ZestFinance for a "System and Method for Building and Validating a Credit Scoring Function"[40] gives an overview of the data sources used for computing the credit score. If an application is submitted through a website, then browser-related behavioural measurements, such as the number of pages viewed by the applicant and the amount of time spent by her on the actual application page, can be used to assess creditworthiness.[41] ZestFinance transforms these measurements into "an ordinal variable on a 0-2 scale, where 0 indicates little or no care during the application process and 2 indicates meticulous attention to detail during the application process."[42]

Similarly, Lenddo, a Singapore based company's patented score predicts an individual's "character or willingness to pay."[43] The LenddoScore complements traditional underwriting tools, like credit scores, by relying exclusively on non-traditional data derived from a customer's social data and online behaviour.[44] An illustrative list of data points considered by LenddoScore includes credit bureau data, telecom, browser, mobile, social networks, e-commerce transaction, financial transaction data, form filing analytics and psychometric data.[45]

While many companies indicate the reliance on alternative data for assessing creditworthiness, the manner and extent of usage of such data for

---

[37] Mikella Hurley & Julius Adebayo, 'Credit Scoring in the Era of Big Data' (2017) 18 Yale JL & Tech 148, 164.

[38] Michael Carney, 'Flush with $20M from Peter Thiel, ZestFinance is measuring credit risk through non-traditional big data' (31 July 2013) <https://pando.com/2013/07/31/flush-with-20m-from-peter-thiel-zestfinance-is-measuring-credit-risk-through-non-traditional-big-data/> accessed 2 January 2020.

[39] ibid. Alya, Guseva & Akos Rona-Tas, 'Consumer Credit Surveillance' in *The Oxford Handbook of Consumption* (1st edn 2019) 351, 352.

[40] US Patent Application 14/276,632, 'System and Method for Building and Validating a Credit Scoring Function', Filed on 13 May 2014, <https://patents.justia.com/patent/20150019405> accessed 3 January 2020.

[41] ibid.

[42] ibid.

[43] 'Our Products, Credit Scoring' (*Lenddo*) <https://lenddo.com/products.html#creditscore> accessed 3 January 2020.

[44] ibid.

[45] Lenddo, 'Credit Scoring Solution', <https://lenddo.com/pdfs/Lenddo_FS_CreditScoring_201705.pdf> accessed 3 January 2020.

determining the creditworthiness of prospective borrowers are still not clear. The opacity associated with such scoring models has evoked new fears and challenges, as discussed below.

III. Challenges associated with Alternative Credit Scoring

The increasing amount of data relied on by alternative data scoring systems raises concerns about the quality and veracity of such data, uncertainty on the use of the data, opacity of scoring methodologies, the ability of such data to predict the creditworthiness of a person, and its potential to discriminate.[46]

Barocas and Selbst note that-

> *"There is no way to directly measure creditworthiness because the very notion of creditworthiness is a function of the particular way the credit industry has constructed the credit-issuing and repayment system. That is, an individual's ability to repay some minimum amount of outstanding debt on a monthly basis is taken to be a non-arbitrary standard by which to determine in advance and all-at-once whether he is worthy of credit."*[47]

Therefore, there is always an element of subjectivity in determining creditworthiness. However, this issue is heightened in the case of alternative credit scoring models that rely on several data points for assessing the creditworthiness of a prospective borrower. The correlation between such data points and the creditworthiness of an individual (including the capacity and willingness of an individual to repay) may not be obvious, and lead to unintended consequences. For instance, if a credit scoring algorithm relies excessively on social media presence, it may be disadvantageous to those without substantial social media presence, who are either denied a loan, or may be required to pay higher interest rates. Further, while a scoring system identifies the factors that are positively or negatively correlated to the ability to repay, it does not analyse the reasons for such a co-relation. Therefore, many characteristics with high predictability may create a disproportionate impact on specific neglected or marginalised classes.[48]

Arguably, to a certain extent, reliance on such big-data scoring tools may reduce subjectivity in loan approvals by decreasing the role of a loan officer's

---

[46] 'Credit Reporting Knowledge Guide 2019' (*World Bank*, 2019) <http://documents.worldbank.org/curated/en/262691559115855583/pdf/Credit-Reporting-Knowledge-Guide-2019.pdf> accessed 10 February 2020; Robinson + Yu, 'Knowing the Score: New Data, Underwriting, and Marketing in the Consumer Credit Marketplace', (October 2014) 21–23 <https://www.upturn.org/static/files/Knowing_the_Score_Oct_2014_v1_1.pdf> accessed 3 January 2020.

[47] Solon Barocas & Andrew D. Selbst, 'Big Data's Disparate Impact' (2016) 104 CALIF L REV 671, 679.

[48] David C. Hsia, 'Credit Scoring and the Equal Credit Opportunity Act' (1978) 30 Hastings LJ 371, 383.

discretion on lending decisions.[49] However, given that such systems employ thousands of data points and complex models, it is often feared that they could potentially be used to mask discriminatory policies.[50] It may also perpetuate existing forms of discrimination if they define "target variables" in a manner that encodes existing bias, rely on inaccurate sample data, or permit the use of proxy variables for sensitive characteristics such as race or gender.[51] Alternative credit scoring algorithms require developers to define the "target variables" and "class labels." Target variable defines the outcomes of interest of the specific task and "class labels" divide all possible values of the target variable into mutually exclusive categories.[52] An example of a class label is creditworthiness, and loan repayment history, details of income, occupations, etc. are examples of target variables.[53] Given that specification of the target variable is not always obvious, developers are required to define the target variables in a manner that can be parsed by computers. This leaves an element of subjectivity in defining such variables, thereby creating the possibility of developers unintentionally parsing a problem in a manner that is disadvantageous to specific classes.[54]

Another concern is associated with the data used to train credit scoring algorithms before they are employed to solve problems. Such training data enables the algorithm to determine correlations between variables and assess its impact on specific classes.[55] However, the presence of bias in the training data may perpetuate past discrimination. For example, a credit scoring model trained on data from a predominantly white population could result in bias against lending to a non-white population.[56] Similarly, if an algorithm draws inferences from a biased sample of the population, "any decision that rests on these inferences may systematically disadvantage those who are under or over represented in the data."[57] Consider a data sample of a class of borrowers that have been historically denied credit for non-commercial reasons such as gender, race, ethnicity or residence in certain areas. Using such sample data for predicting loan default behaviour of underserved borrowers may produce biased results, since such class of borrowers may have been traditionally unfairly excluded from the formal credit system.[58]

---

49 Hurley & Adebayo (n 37) 192.

50 Barocas & Selbst (n 47) 692.

51 Hurley & Adebayo (n 37) 193.

52 Barocas & Selbst (n 47) 678.

53 Vlad A. Hertza, 'Fighting Unfair Classifications in Credit Reporting: Should the United States Adopt GDPR-Inspired Rights in Regulating Consumer Credit?' (2018) 93 NYUL Rev, 1707, 1718.

54 Barocas & Selbst (n 47) 678.

55 Hertza (n 53) 1718.

56 Nikita Aggarwal, 'Law and Autonomous Systems Series: Algorithmic Credit Scoring and the Regulation of Consumer Credit Markets' (1 November 2018) <https://www.law.ox.ac.uk/business-law-blog/blog/2018/11/law-and-autonomous-systems-series-algorithmic-credit-scoring-and> accessed 3 January 2020.

57 Barocas & Selbst (n 47) 680, 681.

58 Majid Bazarbash, 'FinTech in Financial Inclusion Machine Learning Applications in Assessing Credit Risk' (2019) IMF Working Paper WP/19/109, 27 <https://www.imf.org/en/Publications/

When a credit scoring system takes into account multiple data points, it is often feared that the algorithm may indirectly consider sensitive characteristics such as gender, race or religion, even when such data may not have been intentionally fed into the system.[59] A data point can be a genuine criterion for assessing creditworthiness and at the same time, serve as a reliable proxy for sensitive characteristics.[60] For instance, in the US, it has been pointed out that neutral attributes such as zip code can serve as a proxy for sensitive characteristics such as race.[61] Further, consumer's behaviour or habits may vary by race, gender, age, etc.[62] One study showed that thirty per cent of whites use their mobile phone as their sole internet connection compared to roughly 48 per cent of Latinos and 39 per cent of blacks.[63] When these factors are used in credit scoring, such mobile and internet usage could "potentially be used as a proxy for race."[64]

If an algorithm learns that such sensitive characteristics are co-related to credit risk, it is likely to attach higher weightage or significance to such attributes, thereby paving the way for discriminatory selections. Consequently, while a developer may not specifically introduce discriminatory effects into the data mining process, reliance on such data points may put members of a specific class at a disadvantage. Another concern often highlighted relates to "creditworthiness by association."[65] It is likely to happen in cases where the credit scoring algorithm relies on data points that are not solely based on an individuals' data, but based on factors that individual shares with others.[66] For instance, in the US, American Express lowered a customer's credit limit from $10,800 to $3,800. Pertinently, this decision was not based on the person's payment history with the company. It

WP/Issues/2019/05/17/FinTech-in-Financial-Inclusion-Machine-Learning-Applications-in-Assessing-Credit-Risk-46883> accessed 3 January 2020.

59 Hurley & Adebayo (n 37) 182.

60 Barocas & Selbst (n 47) 691.

61 Willy E Rice, 'Race, Gender, "Redlining," and the Discriminatory Access to Loans, Credit, and Insurance: An Historical and Empirical Analysis of Consumers Who Sued Lenders and Insurers in Federal and State Courts, 1950–1995' (1996) 33 SAN DIEGO L REV, 613. This is explained in Hertza (n 53), 1726, where Hertza notes that 'An example of algorithm bias is when credit reports overweigh certain data points, such as consumers' addresses. Consider the harsh reality of racial segregation by neighbourhood and the fact that areas inhabited by a high percentage of racial minorities tend to be less wealthy overall. Areas inhabited by higher-than-average ratios of racial minorities are likely to have lower overall credit scores, thinner credit files, and higher default rates on average. Therefore, negatively scoring particular ZIP codes when calculating an individual's credit score negatively impacts racial minorities.'

62 National Consumer Law Center (NCLC), 'Big Data: A Big Disappointment for Scoring Consumer Credit Risk' (March 2014) 27 <https://www.nclc.org/images/pdf/pr-reports/report-big-data.pdf> accessed 2 January 2020.

63 ibid citing Jessica J Gonzalez, 'Communications Technology Use in Communities of Color: A Presentation to the FCC Technology Transitions Task Force' (18 March 2013) slide 9 <https://transition.fcc.gov/technologies_transitions_policy_task_force/Panel_2-Jessica_Gonzalez.pdf> accessed 2 January 2020.

64 Hurley & Adebayo (n 37) 182.

65 NCLC (n 62).

66 Hurley & Adebayo (n 37) 183.

was based on the fact that other customers of the bank, who had used their card at establishments where that individual recently shopped had a poor repayment history with American Express.[67] In 2015, Facebook secured a patent that may pave the way for lenders to determine a loan application based on the applicant's social networking connections.[68] Briefly, the patent pertains to "authorisation and authentication based on an individual's social network."[69] The patent application indicates that one of the preferred "embodiments of the invention" could be used for credit rating. In the "Summary of the Invention", it is explained that-

> "*When an individual applies for a loan, the lender examines the credit ratings of members of the individual's social network who are connected to the individual through authorised nodes. If the average credit rating of these members is at least a minimum credit score, the lender continues to process the loan application. Otherwise, the loan application is rejected.*"[70]

It is unclear if a product will be developed based on this particular embodiment of the invention. However, critics argue that it could lead to new forms of 'digital redlining.'[71]

Despite the concerns highlighted above, the use of algorithms in alternative credit scoring models may make monitoring such practices difficult. Most of these alternative credit scoring systems are based on machine learning algorithms that operate as 'black boxes' making, it difficult to decipher how such algorithms arrive at a particular decision or make predictions.[72] Such learning algorithms (as opposed to the developer) decide the relevance of data points and attaches weightage to the same.[73] A review of the online terms and conditions of some digital lending platforms in India that employ such alternative credit scoring systems indicates that there is scarce information about how such platforms define creditworthiness or how they select target variables. Further, it has been

---

67 'Facebook is Using You', *The New York Times* (4 February 2012) <https://www.nytimes.com/2012/02/05/opinion/sunday/facebook-is-using-you.html?auth=login-email> accessed 3 January 2020.

68 'Time to reconsider your friends? Facebook patent suggests lenders could check your social connections to decide on whether to approve a loan', *The Daily Mail* (5 August 2015) <https://www.dailymail.co.uk/sciencetech/article-3186001/Time-reconsider-friends-Facebook-patent-suggests-lenders-check-social-connections-decide-approve-loan.html>accessed 3 January 2020.

69 US Patent Application 13/565, 500, 'Authorization and authentication based on an individual's social network' filed on 2 August 2012 <http://patft.uspto.gov/netacgi/nph-Parser?-Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=9,100,400.PN.&OS=PN/9,100,400&RS=PN/9,100,400> accessed 3 January 2020.

70 ibid.

71 Robinson Meyer, 'Could a Bank Deny Your Loan Based on Your Facebook Friends?' *The Atlantic* (25 September 2015) <https://www.theatlantic.com/technology/archive/2015/09/facebooks-new-patent-and-digital-redlining/407287/> accessed 2 January 2020.

72 Matthew A Bruckner, 'The Promise and Perils of Algorithmic Lenders' Use of Big Data' (2018) 93 Chi-Kent L Rev 3, 16.

73 ibid.

pointed out that such credit scoring systems may be protected by intellectual property laws, which also underscores the lack of transparency in which data is collected and used.[74] This implies that consumers may have no recourse to understand how their decisions may impact their credit scores. The lack of transparency regarding the exact data points collected by such systems and how such points are assessed to arrive at a credit score may make it next to impossible for a prospective loan applicant to identify an error in the information that has been collected. Even assuming an applicant could identify such error, it is highly unlikely that she would have the capacity or the expertise to prove that the error resulted in a faulty score.[75] This article does not contend that traditional scoring models are entirely transparent. However, it may be less acute in cases involving such traditional credit scoring by CICs, since these companies and their activities are directly regulated by the Reserve Bank of India ("**RBI**") under the CIC Act.

Access to credit on fair and reasonable terms is critical for persons with fragile financial circumstances. Alternative credit system seeks to capitalise on efficiencies generated by technological innovations such as big data and artificial intelligence to meet such credit needs. However, the challenges associated with these lending models, as discussed in this section indicate the possible negative implications of alternative credit scoring systems on consumers that demand scrutiny by policymakers. As technological innovations continue to influence the availability and the terms of credit provision, they will assume crucial gatekeeping functions in the credit market, determining who can avail credit and on what conditions. As discussed above, the concerns associated with involvement of biased training data for an algorithm, or the usage of sample data that encodes some pre-existing bias, or the ability of an algorithm to combine genuine, neutral data points and treat them as proxies for sensitive characteristics (such as gender, race, etc), may lead to denial of credit to certain groups. A study effect of machine learning on US credit markets notes that "gains from new technology are skewed in favour of racial groups that already enjoy an advantage, while disadvantaged groups are less likely to benefit in this data set."[76] Left unchecked, such challenges bears the risk of encouraging systemic discriminatory behaviour in lending decisions, thereby leading to financial exclusion, especially of such class of borrowers that have historically been denied credit based on the grounds of religion, caste, gender or race etc.

---

[74] See, Brenda Reddix-Smalls, 'Credit Scoring and Trade Secrecy: An Algorithmic Quagmire or How the Lack of Transparency in Complex Financial Models Scuttled the Finance Market, (2011) 12 UC DAVIS BUS LJ 87.

[75] Hurley & Adebayo (n 37) 179.

[76] Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, Ansgar Walther, 'The Effect of Machine Learning on Credit Markets' (11 January 2019) <https://voxeu.org/article/effect-machine-learning-credit-markets> accessed 1 March 2020.

IV. Can The Existing Legal Framework Address These Challenges?

The existing legal framework in India, primarily consisting of the CIC Act, the Information Technology Act, 2000 ("**IT Act**") and the rules issued thereunder, already regulate certain aspects of credit scoring. However, regulators and consumers may find it challenging to apply the existing provisions to alternative credit scoring models due to the nature and volume of data relied on by such systems and the opacity surrounding their operations. This part examines the adequacy of these laws to address the concerns associated with alternative credit scoring systems.

The business of 'credit information'[77] in India is regulated by RBI under the CIC Act. Four companies have been granted certificates of registration as CICs by the RBI under the CIC Act. These are TransUnion CIBIL Limited, Equifax Credit Information Services Private Limited, Experian Credit Information Company of India Private Limited and CRIF High Mark Credit Information Services Private Limited.[78] All 'credit institutions' as defined under the CIC Act (which includes banks and NBFCs[79]) are mandated to become members of all CICs.[80] The CIC Act defines 'credit scoring' as a system "which enables a credit institution to assess the creditworthiness and capacity of a borrower to repay his loan and advances along with discharging his other obligations in respect of credit facility availed or to be availed by him"[81] CICs are authorised to share 'credit information' and provide 'credit scoring' to its 'specified users' or 'specified users' of any other CIC.[82] 'Specified users' refers to any credit institution, CIC, insurance company, telecom service provider, credit rating agency registered with Securities and Exchange Board of India, information utility registered under Insolvency and Bankruptcy Code, 2016, etc.[83] Further, CICs can also provide any person with their 'credit information.'[84] The aforesaid regulatory framework enables credit institutions like banks and NBFCs to obtain credit information about prospective borrowers from a CIC. Further, credit scores computed by CICs is a factor considered by many lenders while granting a loan.

---

77 For definition of 'credit information', see CIC Act, s 2(d). The definition of 'credit information' includes 'the credit worthiness of any borrower of a credit institution.'

78 High Level Task Force on Public Credit Registry for India, Report of the High Level Task Force on Public Credit Registry for India, (April 2018), 7.

79 CIC Act, s 2(f).

80 CIC Act, s 11(1) read with RBI, Membership of Credit Information Companies (CICs), RBI/2014-2015/405 (Notified on 15 January 2015) and RBI, Membership of Credit Information Companies (CICs), RBI/2014-15/458 (Notified on 6 February 2015).

81 CIC Act, s 2(g).

82 CIC Act, s 14.

83 CIC Act, s 2(l).

84 CIC Regulations, 2006 (CIC Regulations), reg 6.

The CIC Act read with the regulations issued thereunder incorporates necessary safeguards for processing 'credit information.' It incorporates several well-recognised privacy principles.[85] For instance, a CIC, credit institution and a specified user must ensure that the data relating to credit information maintained by them is accurate and complete.[86] Such an obligation exists since the processing of incorrect or inaccurate data, can have detrimental consequences for the concerned individual, such as the denial of credit. To ensure meaningful application of such a provision, the CIC Act has imposed corresponding rights on consumers who are providers of such credit information. For instance, the CIC Act expressly provides a framework for such consumers to access their credit information and update or correct such information.[87] The right to access credit information obtained by a credit institution from a CIC is to ensure that the end consumer can understand and gauge the information based on which the financial institution will assess her creditworthiness and her loan application. Given that a CIC will assess the creditworthiness of an applicant based on her credit information, it is essential to ensure that such data is correct and up to date. It will ensure the veracity of decisions based on such data. The CIC Act expressly penalises any unauthorised access to credit information. Further, the CIC Regulations set out the purposes for which such information may be used.[88] It mandates a CIC to adopt privacy principles to collect, process, collate, preserve, share and use credit information obtained via this medium.[89]

The safeguards enshrined under the CIC Act are broadly applicable to CICs, credit institutions or specified users. As discussed in the preceding section many technology companies develop credit risk models and perform a risk assessment for their partner lending institutions, including assessing the creditworthiness of retail borrowers that apply for loans on the platforms developed by such companies. These companies are currently not registered under the CIC Act. Given that these entities provide such services to limited lenders with whom they have a partnership, as opposed to all specified users (as in the case of a CIC), arguably, such entities do not fall within the ambit of the CIC Act. Accordingly, the safeguards enshrined under the CIC Act, including the power of RBI to regulate such credit assessment activity by CICs do not extend to these new-age alternative credit scores. However, data processing by such alternative credit scorers is governed by the provisions of the IT Act and the rules issued thereunder, more particularly the Information Technology (Reasonable Security Practices and Sensitive Personal Data or Information) Rules, 2011 ("**SPDI Rules**"). The SPDI Rules have been issued under Section 43A of the IT Act that provides

[85] These principles are recognised in the recommendations of the AP Shah and the Data Protection Paper.

[86] CIC Act, s 19.

[87] CIC Act, s 21.

[88] CIC Regulations, reg 9.

[89] CICI Act, s 20(1).

for a framework to enact reasonable security practices and procedures for 'sensitive personal data or information',[90] failing which it provides for a compensation mechanism. The SPDI Rules which elaborates on the scope of these reasonable practices and procedures mandates specific requirements for the collection of such information[91] (including provision for consent,[92] purpose limitation[93] and lawful purpose[94]), provision of a privacy policy[95] and also gives individuals the right to correct their information.[96] These rules have often faced criticism for failing to keep up with the pace of technological innovations.[97] The safeguards under the SPDI Rules extend to 'sensitive personal data' which has been defined narrowly.[98] Further, its provisions may be overridden by way of a contract.[99] Moreover, the enforcement mechanism has also been unsatisfactory, with delays in appointment to the post of adjudicators under the IT Act.[100]

A review of the existing framework indicates its inadequacy to mitigate the risks of alternative credit scoring. At the time when the CIC Act and SPDI Rules were enacted, the transparency and confidentiality provisions envisaged therein may arguably have provided an essential safeguard against possible abuses of personal information. However, these measures may not be adequate in the context of alternative credit scoring for reasons discussed herein. First, the framework under the CIC Act is only applicable to credit scoring undertaken by CICs. Currently, technology companies that conduct credit scoring using alternative data are not registered under the CIC Act. The enforcement of the safeguards under the CIC Act has also not been satisfactory, as is evident from recent reports that indicate how credit institutions have indulged in sharing consumer data obtained under the CIC Act with unregulated entities, against the provisions of the Act.[101]

---

90 *See* IT Act, s 43A read with SPDI Rules, r 3.
91 SPDI Rules, r 5.
92 SPDI Rules, r 5(1).
93 SPDI Rules, rr 5(4) and 5(5).
94 SPDI Rules, r 5(2).
95 SPDI Rules, r 4.
96 SPDI Rules, r 5(6).
97 Committee of Experts under the Chairmanship of Justice BN Srikrishna (Justice BN Srikrishna Committee), 'A Free and Fair Digital Economy Protecting Privacy, Empowering Indians' (2018) 7 <https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf> accessed 3 January 2020.
98 Justice BN Srikrishna Committee, 'White Paper of the Committee of Experts on a Data Protection Framework for India' (2017) 17 <https://meity.gov.in/writereaddata/files/white_paper_on_data_protection_in_india_18122017_final_v2.1.pdf> accessed 3 January 2020. The committee notes that definition is limited to include attributes likes medical history, sexual orientation, biometric information and does not cover the wide category of personal data.
99 Ministry of Communications and Information Technology, 'Press Note' (24 August 2011) <https://meity.gov.in/writereaddata/files/PressNote_25811.pdf> accessed 3 January 2020.
100 Justice BN Srikrishna Committee (n 97) 7.
101 'RBI Restricts Access to Credit Data of Consumers', *The Economic Times* (19 September 2019) <https://economictimes.indiatimes.com/small-biz/startups/newsbuzz/apex-bank-restricts-access-to-credit-data-of-consumers/articleshow/71194383.cms?from=mdr> accessed 3 January 2020.

Second, under the existing framework, consumers lack the right to access the data underlying the credit reports, or the right to know what variables are considered by credit scorers to determine creditworthiness or the weightage assigned to such variables. By relying on complex data sets and algorithms to arrive at a credit score, credit scorers are depriving consumers of fully enjoying their rights under the existing framework. Third, to the extent the CIC Act and the SPDI Rules provide a framework for consumers to update or correct their credit information, such safeguards may be outdated and ineffective, given the complexity of algorithms employed by such credit scoring models. Even if a consumer figures out the nuances of how such a credit scoring model works, it is likely to be challenging for her to assess the correctness of a score based on big data and machine learning, which challenges inaccuracies in the raw data, or challenges the insights drawn from the analysis of such data. This complexity is likely to render many safeguards granted under the CIC Act, and the IT Act merely illusory. Fourth, while the CIC Act and the SPDI Rules recognise the principle of purpose limitation, it does not place any restriction on the type of information that may be considered for assessing creditworthiness. Consequently, consumers are neither well informed about the exact data points collected by such scoring models and the basis for a credit decision. Further, unlike in the US[102] that has a specific statutory framework to protect against discrimination in a credit transaction, neither the CIC Act nor the IT Act provides any such protection against possible discriminatory treatment by such scoring models. Notably, while RBI guidelines prohibit discrimination in lending practices,[103] borrowers are likely to find it challenging to prove a case of discrimination in case of alternative credit scoring models that uses multiple data points and complex algorithms.

V. Possible Regulatory Response

The preceding section underscores the inadequacies of the existing legal framework to deal with the possible challenges of alternative credit scoring. The challenges and risks associated with such scoring are closely associated with broader risks of bias in algorithmic decision making that may span across several industries. Therefore, this section examines if the regulatory response to address the challenges in alternative credit scoring can be found in a sector-agnostic law like the proposed data protection law.

To deal with the challenges of the ever-growing digital economy and to give meaningful effect to the right to privacy (that has been recognised as a fundamental right by the Hon'ble Supreme Court),[104] the Personal Data Protection Bill 2018 ('PDP Bill') was introduced in the Parliament in 2019. It contains a

[102] *See* Equal Credit Opportunity Act of 1974.

[103] RBI, Master Circular, 'Loans and Advances – Statutory and other Restrictions', RBI/2015-16/95 (Notified on July 1, 2015).

[104] *See* PDP Bill, Preamble.

comprehensive framework for data protection in India and envisages the establishment of a Data Protection Authority ("**DPA**") to implement the law. The Committee of Experts under the chairmanship of Justice (Retd.) B.N. Srikrishna ("**Justice B.N. Srikrishna Committee**") that was tasked with drafting the law noted the challenges associated with the processing of big data in the context of well-established principles such as collection limitation and purpose limitation.[105] Unlike the CIC Act which applies to credit scoring by CICs, the proposed PDP Bill applies to any 'processing' [106] of 'personal data',[107] including processing that analyses or predicts the behaviour, attributes or interests of a 'data principal' (i.e. the person to whom the data relates).[108] Therefore, the provisions of PDP Bill will also apply to technology companies (which are not CICs) engaged in credit scoring. The broadening of the scope of the legal framework to include all entities that process personal data of consumers will be relevant to ensure fair access to credit.

As discussed earlier, one of the key hurdles faced by consumers in securing their rights under the CIC Act and the SDPI Rules is the lack of information about how their data is being processed. As explained earlier, these concerns are heightened in the case of alternative credit scoring systems. Such systems that rely on machine learning and big data may be too complex for consumers to even comprehend what data is being collected, by whom it is collected and the purpose for such collection and processing. To ensure transparency in processing, the PDP Bill imposes positive obligations on 'data fiduciaries'[109] to notify data principals about the purposes for which personal data is collected, the nature and categories of personal data being collected, the basis for such processing, the procedure for grievance redressal and the existence of a right to file a complaint with the DPA under the law or such other information as may be specified by the regulations.[110] Such provisions may be relevant to ensure that data principals are made aware of various data points that may be collected for credit scoring. It is not clear if the aforesaid provisions may be interpreted to mean that alternative credit scorers will be under an obligation to disclose the manner of computation or weightage

---

105 Justice BN Srikrishna Committee (n 97) 56.

106 PDP Bill, cl 3(31) defines 'processing' in relation to personal data to mean 'an operation or set of operation performed on personal data'. It may include collection, organisation, structuring, use or combination.

107 PDP Bill, cl 3(28) defines 'personal data' to mean 'data about or relating to a natural person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute or any other feature of the identity of such natural person, whether online or offline, or any combination of such features with any other information, and shall include any inference drawn from such data for the purpose of profiling.'

108 PDP Bill, cl 3(14). 'Data principal' has been defined to mean a natural person to whom the personal data relates.

109 PDP Bill, cl 3(13) defines 'data fiduciary' as 'any person, including the State, a company, any juristic entity or any individual who alone or in conjunction with others determines the purpose and means of processing of personal data.'

110 PDP Bill, cl 7(1).

assigned to data variables. However, such information may be required to be disclosed by the DPA through regulations.

The PDP Bill sets out different grounds based on which personal data may be processed, such as consent,[111] compliance with the law, or public function.[112] In addition to these grounds, the PDP Bill also provides for processing on the grounds of 'reasonable purpose'. Explaining the rationale for this ground, the Justice B.N. Srikrishna Committee notes that such a ground for processing provides flexibility to data fiduciaries in situations where they may need to carry out the processing for prevention and detection of unlawful activities including fraud, whistleblowing, and network and information security. The committee notes that in such cases it may not be possible for data fiduciaries to take consent in all situations. Requiring them to seek consent in all cases may "prove burdensome and may raise concerns of consent fatigue among data principals" and hinder the evolution of new technologies based on data analytics.[113] Notably, the PDP Bill defines 'reasonable purposes' to include credit scoring.[114] Consequently, personal data processing for credit scoring purposes will not require the consent of the data principal. The Justice B.N. Srikrishna Committee 'while discussing the rationale for including credit scoring within the definition of 'reasonable purposes'' cites the example of a credit card company that shares personal data of its customers with credit reference agencies for credit scoring. The committee argues that mandating a consent requirement in such cases may not be appropriate if "the credit score of individuals is needed to determine creditworthiness and the real choice is absent for the data principals."[115] One may have wished for a more in-depth discussion by the committee on its rationale for excluding credit scoring from the consent framework. This is particularly relevant in the context of complex alternative credit scoring algorithms and the consequent risks it poses on financial exclusion. Nevertheless, this paper does not stress on consumer consent because ensuring meaningful consumer consent in cases where the consumer and the lending institution does not enjoy the same bargaining power is often doubtful.[116] While consent may not be required for processing data for credit scoring, the PDP Bill confers power on the DPA to lay down safeguards to ensure the protection of the rights of the data principals and determine if the notice provisions as envisaged under the bill will be applicable, "having regard to whether such provision [of notice] would substantially prejudice the reasonable purpose."[117] For credit scoring purposes, this paper argues that the notice requirements, as discussed above, should be retained. This is particularly critical since the PDP Bill dispenses with

[111] PDP Bill, cl 11.
[112] PDP Bill, cl 12.
[113] Justice BN Srikrishna Committee (n 97) 117.
[114] PDP Bill, cl 14(2).
[115] Justice BN Srikrishna Committee (n 97) 118.
[116] Hertza (n 53) 1737–1738.
[117] PDP Bill, cl 14(3).

the consent requirement for credit scoring. By way of regulations, the DPA should expressly set out standardised formats of key information statements that must be necessarily displayed by such credit scoring companies prominently on their websites. This will be in addition to the important terms and conditions that are typically displayed by alternative data scoring companies on their websites. The standardised formats should be designed to draw the attention of the consumer to it before undertaking any transaction with the service provider.

Although PDP Bill dispenses the consent requirement for processing personal data for credit scoring, a data fiduciary must comply with all other obligations under the data protection law.[118] Accordingly, to minimise the concerns associated with big data, the PDP Bill imposes specific obligations on 'data fiduciaries' and confers specific rights on 'data principals'. The PDP Bill not only expands on the principles of purpose limitation,[119] collection limitation[120] and storage limitation,[121] but it also requires data fiduciaries to ensure that personal data processed is complete, accurate, and updated, having regard to the purposes for which it is processed.[122] The PDP Bill requires data fiduciaries to maintain transparency regarding its general practices related to the processing of personal data.[123] The data principal has been conferred with the right to obtain from the data fiduciary confirmation about the processing of personal information, a summary of the personal information processed,[124] and correct and updated personal data.[125] The PDP Bill also recognises the right to data portability[126] that allows data principals to obtain and transfer their data stored with one data fiduciary to another, in a structured, commonly used and machine-readable format. This gives data principals greater control over their personal data and is also likely to foster competition in the sector. Further, the PDP Bill provides *ex-ante* and remedial measures for any 'harm', which includes harm arising from any discriminatory treatment.[127] Remedial measures include the imposition of penalties[128] and compensation.[129]

---

118 Justice BN Srikrishna Committee (n 97) 120.
119 PDP Bill, cl 5.
120 PDP Bill, cl 6.
121 PDP Bill, cl 9.
122 PDP Bill, cl 8.
123 PDP Bill, cl 23. This requires a data fiduciary to make available the following information easily accessible in such form as may be specified by the DPA – categories of personal data collected, the manner of collection, purpose for which personal data is processed and such other information as may be specified by the DPA.
124 PDP Bill, cl 17.
125 PDP Bill, cl 18.
126 PDP Bill, cl 19.
127 PDP Bill, cl 3(20).
128 PDP Bill, cl 63.
129 PDP Bill, cl 64.

The PDP Bill presents a significant avenue for policymakers in India to mitigate the potential risks and challenges presented by alternative credit scoring. Broadly, the principles and safeguards enshrined in the PDP Bill provide the necessary flexibility to regulators and market participants to respond dynamically to these new technological risks. The overarching principles that form the basis of the PDP Bill, if implemented properly, could significantly restrict the potential misuse of personal data of consumers in cases of alternative credit scoring. This approach can be further enhanced through the introduction of more robust standards and safeguards (through the PDP Bill) as outlined herein. Data fiduciaries in case of alternative credit scoring models must be required to make disclosures regarding the data points collected for assessing creditworthiness, the sources from which such data will be collected, collection methods and a description of the weightage given to such data points. Such information must be disclosed to the regulator and consumers periodically and in case of any change in the information. Such a transparency framework will be instrumental to enable consumers to gain a basic understanding of how they are scored. This will be critical for consumers to exercise their rights under the PDP Bill effectively. While the existing framework and the PDP Bill establish accuracy requirements and confer the right on consumers to rectify or update personal information, consumers bear the burden to identify and dispute inaccuracies. As explained earlier, the integration of multiple data points by credit scorers makes it difficult for consumers to verify or dispute the date. This may require shifting the burden of accuracy to the alternative credit's scorers. Besides requiring such scorers to maintain accurate data, the law should also require them to periodically conduct review and audit of their data for accuracy and based on that certify their compliance with the accuracy requirements.

Unlike the European Union General Data Protection Regulation,[130] the PDP Bill does not have a right to object to automated decision making[131] and to access the logic behind it.[132] The Committee of Experts notes that the object of such rights is to curb "harms due to prejudice and discrimination in output data owing to evaluative determinations without human review."[133] The committee noted that this object might be better achieved through an *ex-ante* accountability framework requiring data fiduciaries to set up processes that weed out such discrimination. In this regard, the PDP Bill proposes privacy by design framework,[134]

---

130 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (EU GDPR).

131 See EU GDPR, art 22.

132 See, EU GDPR, arts 13, 14 and 22.

133 Justice BN Srikrishna Committee (n 97) 74.

134 PDP Bill, cl 22. It sets out certain organisational measures for data fiduciaries by setting out certain data handling practices to ensure compliance with the law. These includes policies and measures to ensure that that technical systems are designed in a manner to anticipate, identify and avoid harm to the data principal, technology used in the processing of personal data is in

data protection impact assessment[135] and audit requirements.[136] Notably, such data protection impact assessments and audit requirements apply to 'significant data fiduciaries.' The PDP Bill distinguishes and places additional obligations on 'significant data fiduciaries' which are entities that are capable of causing greater harm to data principals due to their processing activities.[137] A data fiduciary will be notified as a significant data fiduciary by the DPA based on factors such as the volume of personal data processed, the sensitivity of personal data processed, turnover of the data fiduciary, risk of harm, use of new technologies for processing, etc.[138]

Stronger consumer rights as envisaged under the PDP Bill along with the suggestions outlined above could lead to a more competitive credit market by compelling lenders to take into consideration the interests of consumers. Given that the aforesaid framework will make it easier for consumers to port their data to other service providers, the risk of losing out on such consumers may put pressure on lenders to ensure that they comply with consumer requests and adopt fair and transparent procedures in their dealings. While the PDP Bill is a step in the right direction to address some of the challenges associated with alternative credit scoring systems, especially in the context of privacy-related concerns, it may not be sufficient to deal with the potential of discriminatory lending practices by alternative credit scoring systems. This becomes more relevant as algorithmic credit scoring models evolve and become more complicated for consumers to realise their rights under the data protection law meaningfully. In such cases, an algorithmic accountability framework imposing affirmative obligations on developers may be required to deal with the inscrutability and biases of algorithmic decision making. Such a law may provide for the following features. *One,* it should require an impact assessment of the algorithmic credit scoring system, including the training data to assess its impacts on accuracy, fairness or bias. Such an impact assessment should include a description of the algorithm, its design, the training data, the potential risks and measures for mitigating risks. *Second,* it should mandate periodic auditing of algorithms and data collected by it, both at design and implementation stages to check for bias. *Third,* the framework may impose a certification requirement that credit assessment tools have been developed using

accordance with commercially accepted or certified standard and processing of personal data is carried out in a transparent manner.

135 PDP Bill, cl 27. This is *inter alia* applicable for processing involving new technologies or large scale profiling or use of sensitive personal data. The assessment shall include a description of proposed processing operation, assessment of potential harm that may be caused to data principals and measures for mitigating or removing risks.

136 PDP Bill, cl 29. A significant data fiduciary is required to have its policies and its processing of personal data audited annually by an independent auditor.

137 Justice BN Srikrishna Committee (n 97) 160.

138 PDP Bill, cl 26.

empirically sound sample data that does not produce a biased model.[139] The existing framework does not place any substantial limits on the types of data used in credit scoring. Consequently, there is little to prevent scoring tools from inadvertently using data points that may serve as proxies for sensitive attributes such as gender, race or religion. *Fourth*, while minimum standards for conducting an impact assessment or audit should be set out in the legal framework, detailed standards for the same may be left to be developed by industry bodies. *Fifth,* to balance the objectives of promoting innovation and the protection of individual rights, a co-regulatory model of governance may be designed to implement this framework. Under a co-regulatory model, the government and industry share responsibility for drafting and enforcing regulatory standards.[140] The government may set out the objectives and minimum standards in the legal framework, with the implementation being delegated to the industry bodies. To ensure effective enforcement of the framework, the government should have some oversight over the industry bodies through mechanisms such as consultation with the government or approval of the government for drafting standards or codes of conduct, submission of periodic reports to the government, etc. *Fifth*, adequate penalties for non-compliance with the requirements discussed above may be considered. *Sixth*, to ensure that such a framework does not become a hurdle for new and emerging innovative businesses, the framework must adopt a risk-based approach and may restrict its application to specific entities that may be determined based on factors such as turnover, the potential of risk, usage of new technologies, etc.

VI. Conclusion

The advantages and the challenges of alternative credit scoring make it a double-edged sword. While on the one hand, it presents an opportunity to expand access to credit to reach traditionally unserved or underserved consumers, it could also generate new forms of inefficiencies and discrimination. It runs the risk of replicating and in certain cases, even amplifying human biases, particularly in the context of vulnerable groups that have been victims of historical biases. If left unchecked, these algorithms disparate may adversely impact such groups even without the programmer's intention to discriminate.

139 The author acknowledges that this requirement may be easier said than done. Accordingly, the proposed framework should provide for safe harbor provisions that developers may rely on for protecting themselves from any liability as long as they can prove that they have exercised due diligence and care in accordance with the law.

140 Dennis D Hirsch, 'The Law and Policy of Online Privacy: Regulation, Self-Regulation, or Co-Regulation?' (2011) 34 Seattle University Law Review 439, 441 <https://digitalcommons.law.seattleu.edu/cgi/viewcontent.cgi?article=2003&context=sulr> accessed 3 January 2020; Hans-Bredow-Institut and Institute of European Media Law, 'Final Report: Study on Co-Regulation Measures in the Media Sector' (June 2006) 17 <https://www.hans-bredown institut.de/uploads/media/default/cms/media/cd368d1feeoeocee4d50061f335e56291846i245.pdf> accessed 9 April 2018.

Tapping the potential of algorithmic lending and containing its possible risks calls for reassessing the existing regulatory framework. The preceding section already outlines possible regulatory responses. The measures in the suggested framework described above can be viewed as a medium to long term recommendations. Till the time such an appropriate regulatory response to the risks of credit scoring is implemented by policymakers, the RBI's supervisory authority over CICs, credit institutions (banks, NBFCs, peer to peer lenders) should be leveraged to gain insights on how to best promote the promise of algorithmic lending while preventing harmful credit discrimination. Such insights will be instrumental in designing safeguards and standards under the PDP Bill and the algorithmic accountability framework, as discussed above. Further, RBI may also exercise its extensive powers to issue directions to entities within the scope of the CIC Act to ensure that the alternative credit scoring models used by them while providing loans are transparent and do not indulge in arbitrary or discriminatory decisions.[141]

The use of alternative data and machine learning in assessing creditworthiness presents a unique opportunity to bring the unbanked and the traditionally underserved consumers within the fold of the formal lending ecosystem. From a lenders perspective, such systems provide a means to increase the efficiency of the loan underwriting process. However, as with most algorithmic decision making, such credit scoring models also pose risks for credit markets. Given the consequences that can flow from a faulty or unfair credit score, there is a need to ensure that developers and users of such credit scoring models tread cautiously. This article calls for caution in the use of such alternative credit scoring models through a regulatory framework that seeks to ensure transparent, accurate and fair credit scoring decisions and at the same time does not unduly inhibit such innovative technologies.

141 CIC Act, s 11.