



THE INDIAN JUSTICE SYSTEM SIGNIFICANTLY IMPROVED BY USING BIG DATA UTILIZATION AND ARTIFICIAL INTELLIGENCE TECHNIQUES

¹Naga Vijay Kumar, ²Kotaru Kumar Ashwin, ³Dr. Swathirao Suguru, ⁴Madduri Nikhitha
³Professor, ^{1,2}Assistant Professor, ⁴Student, ^{1,2,3,4}Department of Computer Science Engineering ,
Siddhartha Institute of Engineering and Technology, Hyderabad, India.

ABSTRACT

In India, policy development is increasingly focusing on artificial intelligence (AI). Regardless of where the reader of this article lives, the country's geographical significance, expanding AI sector, and ambitious governmental activities surrounding AI make it a crucial jurisdiction to consider. The constraints and hazards of data-driven judgments are still taken into account in India and numerous other countries when developing and deploying AI systems, despite the fact that current policy procedures aim to promote the fast development of AI for economic growth and social benefit. This article makes the case that the sociological and ethical issues raised by the technological constraints of AI systems should be taken into consideration when designing policies to inform what policy processes aspire to achieve. It proposes a framework for such deliberation to occur, by analyzing the three main stages of bringing machine learning to deployment—the data, model and application stage. It is written against the backdrop of India's current AI policy landscape, and applies the proposed framework to ongoing sectoral challenges in India. With a view to influence existing policy deliberation in the country, it focuses on potential risks that arise from data-driven decisions in general, and in the Indian context in particular.

INTRODUCTION

John McCarthy, the pioneer of artificial intelligence (AI), initially presented the concept at Dartmouth in 1956. Risks come with digital transformation since technology is the foundation. AI and machine learning (ML)-based technologies have steadily improved in capabilities and accessibility over the past few years, and this trend is not expected to stop. Understanding the law of AI in the future, as well as its benefits and drawbacks, can help mankind decide whether or not to use AI. The goal of AI research and regulation is to strike a balance between social security benefits of innovation and possible drawbacks and barriers. The Indian government has prioritized the development, implementation, and promotion of AI in order to improve society's quality of life. The specificity and authenticity of the information regarding the locations of the crimes, together with specifics on how they are represented, gave an approach to understanding such crimes in other countries. McGuire and Holt's further throws light on the impressive and much needed Routledge Handbook of Technology, Crime and Justice that has evidence of criminology's burgeoning of technological interest. The most important lookout to implement this research would be to update judges to be specialist in the field of computer; such laws should be implemented wherein all the judges should be well trained to use this technology.

Let's understand the meaning of Artificial Intelligence first. What is Artificial Intelligence? Artificial Intelligence [AI] is also an automatic processing system able to perform tasks that ordinarily require human intelligence. Many of these AI systems are powered by machine learning, variety of them are powered by deep learning and some of them are powered by very boring things like rules. This comes with learning which involves garnering the principles and knowledge for using the

information. All thanks to data-based service industries it's becoming very popular and necessity. As per the father of computing, —The science and engineering of constructing intelligent machines, especially intelligent computer programs|| 1. Artificial Intelligence can be how of construct a computer, a computer- controlled robot, or software who think intelligently, within the similar manner as the intelligent humans think.

AI is accomplished by studying how human brain thinks, and therefore the way humans learn, decide, and work while trying to resolve a problem, so using the outcomes of this study as a basis of developing intelligent software and systems. Philosophy of AI While exploiting the power of the pc systems, the curiosity of human, lead him to wonder, —Can a machine think and behave like humans do?|| Thus, the event of AI started with the intention of constructing similar intelligence in machines that we discover and regard high in humans.

Preparing the Model

First, the data is loaded in python and then we perform data cleaning and exploring the information in the variables. Pandas which provide data frames are imported using python, Matplotlib provides plotting support, and Numpy provides scientific computing within dimensional object support as seen in Fig. (2).

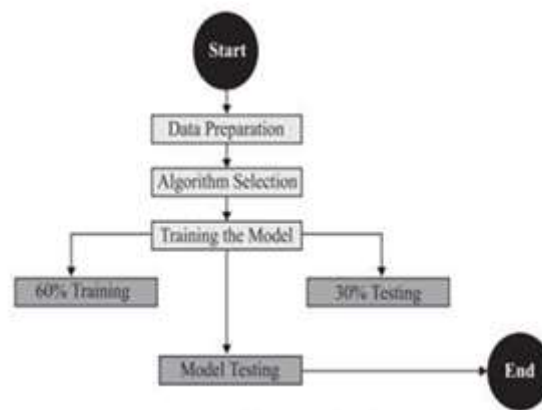


Fig. 1. Model Process Flowchart

```

In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as seabornInstance
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics
import os
%matplotlib inline

# pandas is a dataframe library
# numpy provides N-dim object support
# matplotlib.pyplot plots data
  
```

Fig. 2. Importing Libraries

Secondly, standardization and visualization of data is very important to ensure that data fits the assumptions of the models. The Universal Rule of Law states that human rights, democracy and development depend on the level of progress the organizations and governments can achieve on the criminal justice front. The primary and crucial objectives of the criminal justice are controlling and preventing crime, maintaining law and order, protecting fundamental rights of victims along with the people in conflict with law, punishment and rehabilitation of those adjudged guilty of committing of crimes, and protection of life and property against crime and criminality in general. It is considered to be the primary obligation of the state under the constitution of India.

This paper would thus give an overview how every police station can update their data and predict the criminal behavior of the crime or any data available. Importing various libraries and functions is the positive point of using python in this research paper since the data could be easily adjusted. Accurately predicting rare events is difficult, so the probability of having them in data is low, and the probability of training the algorithm is also low. Therefore, we only need a few percentages of the

event to be able to train, to ensure that we have a reasonable chance to define how correctly a person or state is likely to develop the behavior or motive of committing a crime. Importing pandas will let us easily search the columns by name and see how many times this is true. When a crime is predicted there will be questions arise regarding how an algorithm or code can be trustworthy. This research would, therefore, throw light on this area where the data itself would be deciding everything, the more real the data the more effective the accuracy would be. Data mining and predictive analysis play an essential role in our life. Now if we look into the data available very carefully, we can find whichever states having high unemployment rate. It is noteworthy, that such states have high cybercrime rates which further denotes that in various states computer is used as a source to dupe money through various online frauds. The reason behind this is maintaining the anonymity and causing the harm because of vengeance or other motives. Cybercriminals mostly exploit the high-speed internet available at a lower cost to commit various criminal activities without being caught unless the states possess properly well-maintained cyber security labs to curb such crimes. The CMIE report further reveals that people belonging to age group 40 to 59 years have been successfully able to retain their jobs whereas people aged below 40 years were expelled out of their respective jobs which lead to social tension, desire of revenge, anger and other motives to launch such cyber-attacks.

METHODOLOGY

Having discussed the policy landscape in India, I now turn to the technical considerations that can help inform policy development. This section introduces a novel framework to understand the limitations of data-driven decision-making processes, with particular reference to the Indian context. It will do so by using the technical template of a machine learning system in three stages, i.e. data, models and application. While some considerations discussed bleed into different stages, this taxonomy emphasizes the most appropriate point at which safeguards should be put in place. Data Machine learning depends on data in order to improve performance on future tasks.

Existing scholarship has shown that bias in data can lead to disproportionately adverse outcomes that only entrench discrimination that exists in society [6]. The approach of identifying, assessing, and mitigating bias, as suggested in the NITI Aayog report, is one that is intuitive and simplistic at the same time. This is because bias arises not only from how data are collected, but also from the environment in which it was generated. Training data, thus, raise the following issues.

Access to data

At the risk of using a be-laboured metaphor, data are the oil that fuels AI. Available, accessible, accurate and affordable data are key to building AI systems. This is often a challenge in India. Data that are both accurate and relevant in a given context are rarely readily available. Crime data in India, for example, are grossly underreported [7]. This constraint is felt by burgeoning FinTech and healthcare-related start-ups in the country; very few have access to accurate, affordable data about the populations that they hope to serve. Ryan Calo terms this the problem of data parity [8], where only a few well-established leaders in the field have the ability to acquire data and build datasets. India's National Data Sharing and Accessibility Policy contemplates sharing of non-sensitive data generated using public funds through the Open Data Platform, but this has had moderate success in solving the data parity problem, as a majority of quality data in India is restricted solely to the private sector [9].

Bias in collection

Datasets used to train models run the risk of having 'dark spots' [1] where communities and classes of individuals are overlooked. Data from minority and disadvantaged groups may be more difficult to collect, access or verify. Machine learning depends on generalizing based on examples. If the examples used to train a machine learning system under-represent certain groups that live on the margins [1] of data collection, the generalization offered will discriminate against the under or over-represented group, depending on the specific case. This is especially dangerous in jurisdictions like

India, where being able to have a data footprint is a function of privilege in the first place: either due to gender, caste, geographical location or class [2]. And yet, to proactively include individuals into datasets that they would have otherwise been left out of for the purpose of accuracy in learning models is effectively assuming that the relevance of technologies that aid in profiling and surveillance of those individuals is a given [3]. This problem is only amplified when we consider that surveillance is never neutral, it is fundamentally disproportionate in context of gender, caste, race and religion

Systemic and historical bias

Data imbibe historical discrimination that exists in the environment in which it was created. This leads to a third layer of complication. Often, systemic bias in data is considered problematic because of the outcomes at the time of decision-making. For instance, take FaceTagr, a facial recognition app currently being used by policeman in Chennai, Tamil Nadu. FaceTagr enables policeman to photograph individuals who ‘look suspicious’ [5] and query these photos against existing criminal databases. As Bhatia observes, this is inherently problematic because, ‘in the eyes of the police, —a guy who —looks suspicious for walking on the road at 2 AM will invariably come from a certain socio-economic class, measured simply by his appearance. And in this way, the FaceTagr app will invariably facilitate targeting the homeless and the destitute.’ [6]. However, this is only one class of problems that arises from historical and systemic bias. Not all problematic cases lead to adverse outcomes that can be quantified; some are so because they entrench and sometimes exacerbate real-world sentiments of racism, sexism, violence and hate. For instance, let us consider Google’s AI-powered search algorithm. A Google search for ‘south Indian masala’ returns results not of spices, but of women.¹ This is not necessarily a flaw in the algorithm, but an uncomfortable reflection of society’s stereotypes. Data can thus not only expose problematic stereotypes and bias, but also cement, formalize and imbibe it, simply by virtue of representing the environment in which it was created.

LITERATURE REVIEW

Massive amount of data is created, acquired and stored every day because of organizations’ daily activities with the application of new technologies and systems. It is a challenge to deal with it efficiently and to get an understanding of it. As reminded by Lyko et al. (2016) most data is of potentially high value. However, if there is no Big Data analysis, many of these data gathered would have no use. Among these organizations is the public sector that, according to Munné (2016), starts to be conscious of the possibility to obtain value from Big Data. It is not different with the Judiciary in Brazil, which is nowadays using the Electronic Judicial Process (PJe) to carry out its judicial activities.

The use of PJe produces large quantities of data that can be invaluable for many legal actors, however, this information is very complex and diverse. The complexity lies in the creation, source, structure, management, handling, storage, integration and utilization of this knowledge. These data can be employed for judicial decision-making with the application of Big Data technologies and analysis, which might increase the efficiency of the Brazilian judiciary system, bringing satisfaction to magistrates, saving public expenditures, causing a positive impression on citizens and on the society.

Big Data is an area with vast applicability such as biomedical, healthcare, environmental, business, management, government, business analysis, product development, tourism, marketing, transportation, education, national security, criminology, finance, insurance, telecom, media, entertainment, retail, manufacturing, weather information, science, energy, smart cities, and disaster management.

There are many definitions of Big Data in the literature, most of them agree about the fact that Big Data can help improving the quality and value of decision-making. In addition, it usually involves complex, large, diverse, evolving and increasingly huge volume of structured, semi-structured and unstructured data from various traditional and modern sources. —Big Data is the emerging

field where innovative technology offers new ways to extract value from the tsunami of available information.¶

Nevertheless, the consensus is seen in the characteristics of Big Data, the 3 Vs – volume, velocity and variability. Volume stands for the huge size of the dataset, velocity is related to the fast speed data is generated and used as well as its dynamic features, while variability indicates the vast number and plurality of sources available. Veracity and value are two other Vs incorporated to Big Data characteristics. Veracity represents the trust and reliability of sources as well as the accuracy, quality and certainty of data. Value refers to the worth, usefulness and importance of Big Data to an organization in terms of knowledge creation and decision-making, it may be specified in terms of financial gain or savings. Value can be created "by acquiring data, combining data from different sources and across sectors, providing low latency access, improving data quality, ensuring data integrity, enriching data, extracting insights, and preserving privacy¶. There is a characteristic that is becoming increasingly important – valence – and it is related to the inter- connectedness of the data points which expands fast when more data and data sources are added. Ghosh and Nath (2016) affirmed that —data items are often directly connected to one another¶ and —valence measures the ratio of actually connected data items to the possible number of connections that could occur within the collection.¶

The large volume of data has brought some challenges and opportunities to organizations. Opportunities suggested are automation of decision-making process with computational access to available data and knowledge searching process; intelligent decision and forecasting; information sharing covering various domains and institutions; decreasing time for data-driven decision and recommendation using past, historical and current data through simulation, modeling, exploitation, comparison as well as predictive and visual analytics; increasing productivity in public sector that has important budgetary limitations; facility to obtain information quickly; generation of predictions, prescriptions and recommendations.

CONCLUSION

The application and development of computing technology within the judicial field are in no time. it's predictable that the intelligence technology are used more and more widely in cases where the facts are fully proved and easier to type, even in difficult and sophisticated cases, the new trial mode of —human-computer collaboration¶ could also be realized. it's mainly supported the neural network and machine depth learning theory breakthrough. Improving the applying efficiency of technology within the judicial trial field is that the direction of our future efforts. The contribution of this text is that it tries to elucidate the whole process of the judicial application of computer science supported close integration into the standard judicial pattern. This study pays close attention to the event of existing theoretical research and tries to settle on the optimal operating framework. ranging from the speculation of judicial computing and therefore the regular judicial pattern of Chinese litigation procedure's principle and proof mechanism, the paper analyses the idea and practice of AI applied within the judicial field supported the interior and external factors affecting judicial decisions. The article seeks to demonstrate the function, functional expectations, limitations and risks of AI within the litigation field. additionally, it analyses the legal aspects of computing, the moral issues generated by the event of computing, the response stimulus, the thorough integration bottlenecks of computing and legal reasoning, and possible solutions to resolve problems.

This article is principally about a summary of phenomenon interpretation and general rules, that is, the judicial application process of computer science is discussed from the macro perspective. However, so as to comprehend the successful operation of computing within the judicial field, the applying of AI technology within the judicial field has to be studied from the macro-level. Therefore, the deficiency of this text lies within the lack of necessary thinking on a way to make sure the implementation and do the concept of fair trial within the process of application when it faces very detailed and important problems, like a way to achieve the balance between —fairness¶ and —efficiency¶ within the process of application.

REFERENCES

1. Balkin J. 2017 Free speech in the algorithmic society: big data, private governance, and new school speech regulation. U.C. Davis L. Review, Forthcoming 2018. See https://papers.ssrn.com/sol3/paperscfm?abstract_id=3038939.
2. Angwin J, Larson J, Mattu S, Kirchner L. 2016 Machine Bias. ProPublica. See <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
3. O'Neil C. 2016 Weapons of math destruction: how big data increases inequality and threatens democracy, pp. 278–315. New York, NY: Crown Publishing Group.
4. Liping Z. 2018 China is using AI to tackle traffic jams. World Economic Forum. See <https://www.weforum.org/agenda/2018/01/china-is-using-ai-to-tackle-traffic-jams>.
5. Lynch S. 2017 Andrew NG: Why AI is the new electricity. Stanford Graduate School of Business. See <https://www.gsb.stanford.edu/insights/andre-w-ng-why-ai-new-electricity>.
6. NITI Aayog. 2018 National strategy for artificial intelligence. Niti Aayog46. See http://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AIDiscussion-Paper.pdf.
7. Gillespie T. 2014 The relevance of algorithms, Media technologies: essays on communication, materiality and society 167.
8. MacCormick J. 2012 Nine algorithms that changed the future, p. 3. Princeton, NJ: Princeton University Press.
9. Elish MC, Hwang T. 2016 An AI Pattern Language, Intelligence and Autonomy initiative (I&A) at Data & Society. See https://www.datasociety.net/pubs/ia/AI_Pattern_Language.pdf.
10. Munné, R. (2016) Big Data in the Public Sector. In: New Horizons for a Data-Driven Economy. Springer, Cham, [e-book]. https://doi.org/10.1007/978-3-319-21569-3_11
11. Floridi L. 2016 Should we be afraid of AI? Aeon. See <https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible>.
12. Surden S. 2014 Machine learning and the law. Washington Law Review 89, 87– 115.
13. Munoz A. 2014 Machine learning and optimisation. New York University. See https://cims.nyu.edu/~munoz/files/ml_optimization.pdf.
14. Ghosh, K., Nath, A. (2016) Big Data: Security Issues and Challenges. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), [e-journal] 3(3), pp.1-9. <http://dx.doi.org/10.20431/2349-4859.0303001>
15. Lyko, K., Nitzschke, M., Ngonga Ngomo, A. C. (2016) Big Data Acquisition. In: New Horizons for a Data-Driven Economy. Springer, Cham, [e-book]. https://doi.org/10.1007/978-3-319-21569-3_4.