

## **PREDICTIVE MODELLING TECHNIQUES AND APPLICATIONS**

Saanvi Arora

Delhi Public School

DOI: 10.46609/IJSSER.2021.v06i12.024 URL: <https://doi.org/10.46609/IJSSER.2021.v06i12.024>

Received: 17 Nov. 2021 / Accepted: 5 Dec. 2021 / Published: 31 Dec. 2021

### **ABSTRACT**

Predictive modelling has developed rapidly in the last decade and is now used in every industry, with predictive modelling algorithms governing almost every area of life from consumer industry to healthcare and transportation. From more basic techniques, predictive modelling has evolved into more complex structures such as neural networks, increasingly using artificial intelligence and machine learning. While acknowledging the benefits and ease these techniques have brought to society and increased innovation, this paper argues that the challenges and social implications of predictive modelling cannot be ignored. Increasing concerns regarding the lack of accountability of algorithms, unintended effects on human rights and opportunities as well as privacy concerns must be addressed for predictive modelling to be harnessed in a way that is transparent and serves the well-being of the greatest number of people. This paper examines these key arenas of challenges through case studies and proposes policy interventions to improve the framework of accountability for corporations and governments which make use of predictive modelling algorithms and decrease bias in the datasets which these algorithms are trained on.

### **Introduction**

Predictive modelling has emerged as a tool and method of importance not only in scientific disciplines but also for industrial and commercial real-world applications, due to its often attributed high accuracy levels. Predictive modelling is a statistical technique which utilizes machine learning and data mining to predict future outcomes with the aid of historical and existing data sets (Ali, 2020). By analyzing both current as well as historical data, predictive modelling is used to predict a variety of outcomes in a myriad of real world applications, from predicting TV ratings, to weather and disease forecasts, to usage in the stock market, insurance and online advertising and marketing, content curation, virtual assistants and so on (Ali, 2020; Omnisci, n.d; Bhatt, 2021).

Some types of popular predictive modelling techniques and categories include logistic regression, decision trees, time series analysis, classification models, clustering models, forecast models, and outliers models (Ali, 2020; Omnisci, n.d). With interactive and easy-to-use software becoming more prevalent, predictive analytics is no longer just the domain of mathematicians and statisticians. Business analysts and line-of-business experts are using these technologies as well (SAS, n.d). With increased digitization and algorithm driven business models, predictive analysis is emerging as the foremost method to drive business operations, understand market trends, assess risk, and even increase security and prevent fraud (SAS, n.d).

However, predictive modelling is not foolproof and has raised concerns due to societal implications of its past failures. For example, predictive modelling is a methodology that has been widely used in the financial industry in the past and some of the major failures contributed to the financial crisis of 2007–2008 (Du, 2020). These failures exemplify the danger of relying exclusively on models that are essentially backward looking in nature. Moreover, predictive modelling has raised concerns with respect to bias in data sets which have disproportionately affected racial and gender minorities and have excluded people from opportunities for employment and financing in extreme situations (Du, 2020). Furthermore, concerns have been raised regarding the potential for predictive models used in targeted advertising and social media algorithms to create echo chambers, increase the proliferation of fake news, create privacy concerns due to collection of personal information, and lead to behavioral modification.

This paper will examine the development of predictive modelling techniques and describe the primary types of techniques and categories. The paper will also examine the real world applications of predictive modelling and emerging technologies in this space. The paper will then critically analyze the challenges, limitations and social implications of predictive modelling and pose policy recommendations for addressing the future challenges of predictive modelling and more complex techniques which are emerging.

## **Background**

In the past decade, predictive modelling has seen growth from a niche method used by a select few companies which possessed the resources to technique used by the majority of private and public sector enterprises for all manor of analysis. Today, there are a variety of predictive data models that have been developed to meet specific requirements and applications (Selerity, n.d). With the rise of algorithm and data driven business models in all arenas from social media to video recommendations on YouTube to e-commerce related personalized recommendations, machine learning and computational models have seen a dramatic surge. Machine Learning

allows you to make products that adapt to consumer's needs based on predictive modeling. The program would be provided a collection of data sets and training instances for the consumption of the machine so that they might create models and predictions based on these sets of instances (Du, 2020). Various methods of predictive modelling have developed over time. Linear regression analysis is a predictive modelling technique that estimates the relationship between two or more variables, and specifically allows one to examine how the typical value of the dependent variable changes when any one of the independent variables is varied (Panchotia, 2020).

For example, linear regression has a common real world application in finance and sales, for example predicting sales of a company based on estimates of credit card expenditure (Panchotia, 2020). Among the most commonly used techniques is classification analysis, which categorises data for a simple and direct query response, which is in turn determined by categorising information based on historical data (Selerity, n.d; Ali, 2020). Classification models are used in nearly every industry due to its broad applications and ease of retraining the data, and can therefore predict user and customer trends in retail and subscription models, plan marketing campaigns to specific customers, content recommendation, improving revenue cycle management, and even understanding diseases by providing an accurate diagnosis based on past data (Bhatt, 2021; Parthasarathy, 2021).

Among the more complex techniques are neural networks, which has seen a sharp increase in adoption and consequent research and development efforts in the past decade. A neural network is a predictive modelling technique inspired by the working of the human brain, to process past and current data to estimate future values. Neural networks tend to have high accuracy even if the data has a significant amount of noise (Bari et al, n.d). Neural networks are used to find correlations in exceptionally large data sets and "to learn" and identify patterns within the data (Ali, 2020; Warudkar, 2020). Neural networks can also be used to solve other problems like the human brain, through artificial intelligence and machine learning, which provides applications in facial recognition technology, voice and speech recognition, image detection, self driving vehicles among other complex technologies (Warudkar, 2020). Unlike linear regression and classification analysis, neural networks use a hidden layer of data to make predictions more accurate and 'learn' in the way the human brain does, being able to cluster data, recognize similarities and allow the algorithm to learn on its own (Warudkar, 2020). The following sections of this paper will discuss the social implications of these emerging technologies and the challenges of the continually developing field of predictive modelling, as well as implications for policy.

## **Discussion**

With the rapid development of more complex predictive modelling techniques, and techniques such as neural networks which learn unsupervised by human intervention, concern has been mounting about the social implications of these techniques. Particularly, as algorithms are increasingly used to predict every aspect of every day lives and make predictions about people with respect to important aspects such as employment, financing and incarceration, the social issues of predictive modelling have come to the fore (Knowledge@Wharton, 2019). The potential problems arise due to fact that predictive modeling algorithms are trained on the basis of biased data which is found in the real world. This has led to consequences such as racial bias in predicting recidivism, gender bias in employment algorithms, therefore, perpetuating the structures of oppression of the real world (Knowledge@Wharton, 2019). Predictive modelling techniques used for targeted advertising has also been known to create biased outcomes. For example, Facebook recently faced a lawsuit for violating housing discrimination laws as the algorithm allowed for landlords to exclude people of a certain religion, gender, or race from seeing their advertisements for housing (Du, 2020). Errors in data labelling, shortages in datasets required to train the algorithm, and mistakes made in generalizing machine learning (Ali, 2020), are all key challenges to contemplate, especially due to the real adverse effect that algorithms can have on human rights.

Therefore, it is crucial to address the importance of the validity of mainstream statistical and machine learning techniques as the accuracy, transparency, and neutrality of the initial dataset can have serious implications to all future decisions made. Considering the example of predictive modelling in the criminal justice system, a study conducted in 2016 by ProPublica discovered that the algorithm underestimated the likelihood that white defendants would re-offend but overestimated the likelihood for Black defendants, as the algorithm was trained on historically biased datasets (Fox, n.d). Since African Americans have historically been arrested at higher rates due to racial inequality, algorithms for recidivism trained on biased data produced biased outcomes (Fox, n.d). Similarly in 2014, Amazon experimented using a predictive modelling based software to screen potential applicants for jobs. The software preferred male candidates, due to being trained on a decade of resumes which were majority male (Fox, n.d). A large area of concern is predictive modelling and machine learning in social media algorithms, which are increasingly determining what views, news and opinions users are subjected to which has led to proliferation of fake news and even violence in some countries, as well as has huge implications for manipulation of elections and voting decisions (Lau and Akkaraju, 2019).

Machine learning algorithms used for content moderation has also created impacts such as shadow banning, which impacts the kind of communities which can put out content, make an income selling their goods and services and express their opinions on important issues (Brown, 2021). For instance, even though Facebook/Instagram says they allow non-pornographic nudity in fine art posts, an algorithm is still likely to flag art featuring nudity depicting dark-skinned bodies, fat bodies, female bodies, or LGBTQ or BIPOC culture and content (Brown, 2021). The algorithm is then triggered to push such users' content out of the feed. Such action further marginalizes already marginalized groups, making those who are underserved and ostracized in society even more invisible and underserved on social media (Brown, 2021). In addition, privacy and ownership of data is a key consideration and concern for predictive modelling techniques used in applications such as facial recognition and collection of personal data without consent, as well as how this data and stored and sold (Du, 2020). Predictive modelling in dating apps reinforce bias as well. A review by researchers at Cornell University identified similar design features for some of the same dating apps — and their algorithms' potential for introducing more subtle forms of bias (Lau and Akkaraju, 2019). Another study found that most dating apps employ algorithms that generate matches based on users' past personal preferences, and the matching history of people who are similar (Lau and Akkaraju, 2019). This has the effect of in fact a subtle form of segregation.

Algorithmic accountability is the need of the hour, to ensure that AI does not use historical data to pre-judge outcomes; implemented incorrectly, AI will only repeat the mistakes of the past. It is imperative that data and computational scientists integrate input from experts of other domains, such as behavioral economics, sociology, cognitive science, and human-centered design, to create algorithms which are transparent, ethical and are created based on principles of fairness (Lau and Akkaraju, 2019). It is further important that the right to privacy of users from whom data is collected to train these algorithms is respected.

## **Conclusion**

Predictive modelling is a continually developing space with a number of real world applications which seem to permeate into every area of life. The challenges of accurate data collection and training of algorithms are crucial to address given the impact of predictive modelling on the rights of people, including exercising their right to vote, freedom of expression, access to credit and insurance, and employment. Regardless of the stakes, it must become an indispensable part of the future development of predictive modeling and especially Artificial Intelligence based modelling, adopt universal guidelines for creating and using AI, such as democratic values, afe, open, and obvious to users, while those who make and use AI should be held responsible for

their actions and offer transparency (Lau and Akkaraju, 2019). Achieving parity and validity in the initial dataset for all variables has significant implications for privacy and personal liberties. Therefore, it is crucial that the predictions of the algorithm do not differ by indicators such as race, sex, class, and so on (Knowledge@Wharton, 2019).

It is also important for users to develop a sense of awareness about the impacts of predictive modelling on their daily lives, in one part through self action and in the other part through transparency from those who own the data. Predictive modelling has numerous benefits and applications. However, policy makers and larger society must consider the extremes and impacts on human rights when predictions can be incorrect, approaching all modelling with a sense of transparency by design.

## References

Ali, R, (2020), 'Predictive Modeling: Types, Benefits, and Algorithms', Netsuite, <https://www.netsuite.com/portal/resource/articles/financial-management/predictive-modeling.shtml#target7>

Bari, A, (n.d), 'How Predictive Analysis Neural Networks Work', Dummies, <https://www.dummies.com/programming/big-data/data-science/how-predictive-analysis-neural-networks-work/>

Brown, A, (2021), 'Understanding The Technical And Societal Relationship Between Shadowbanning And Algorithmic Bias', Forbes, <https://www.forbes.com/sites/anniebrown/2021/10/27/understanding-the-technical-and-societal-relationship-between-shadowbanning-and-algorithmic-bias/?sh=2adcfccb6296>

Du, M, (2020), 'Modern Challenges of Predictive Modelling', Medium, <https://medium.com/swlh/modern-challenges-of-predictive-modelling-74c273b290b6>

Fox, P, (n.d), 'Bias in predictive algorithms', Khan Academy, <https://www.khanacademy.org/computing/ap-computer-science-principles/data-analysis-101/x2d2f703b37b450a3:machine-learning-and-bias/a/bias-in-predictive-algorithms>

Knowledge@Wharton, (2019), 'Removing Human Bias from Predictive Modeling', <https://knowledge.wharton.upenn.edu/article/removing-bias-from-predictive-modeling/>

Lau, T, and Akkaraju, U, (2019), 'When Algorithms Decide Whose Voices Will Be Heard', Harvard Business Review, <https://hbr.org/2019/11/when-algorithms-decide-whose-voice-will-be-heard>

Omnisici, (n.d), 'Technical Glossary: Predictive Modelling',<https://www.omnisci.com/technical-glossary/predictive-modeling>

Panchotia, R, 'Predictive Modelling Using Linear Regression', Medium,<https://medium.com/swlh/predictive-modelling-using-linear-regression-e0e399dc4745>

Parthasarathy, S, (2021), 'What is Predictive Analytics?',<https://www.logianalytics.com/predictive-analytics/what-is-predictive-analytics/>

SAS, (n.d), 'Predictive Modeling Techniques',[https://www.sas.com/ko\\_kr/insights/analytics/predictive-modeling-techniques.html](https://www.sas.com/ko_kr/insights/analytics/predictive-modeling-techniques.html)

Selerity, (n.d), 'Types of predictive analytics models and how they work',<https://seleritysas.com/blog/2019/12/12/types-of-predictive-analytics-models-and-how-they-work/>

Warudkar, H, (2020), 'Prediction using Neural Networks', Express Analytics,<https://expressanalytics.com/blog/neural-networks-prediction/>