



## **Fair use defences during copyright litigation: is the success of a fair use defence strategy predictable?**

D'Rosario, Michael. 2017. Fair use defences during copyright litigation: is the success of a fair use defence strategy predictable? *International journal of strategic decision sciences*, vol. 8, no. 2, pp. 1-21.

DOI: [10.4018/IJSDS.2017040103](https://doi.org/10.4018/IJSDS.2017040103)

©2017, IGI Global

Reproduced with permission.

Downloaded from DRO:

<http://hdl.handle.net/10536/DRO/DU:30109310>

# Fair Use Defences During Copyright Litigation: Is the Success of a Fair Use Defence Strategy Predictable?

Michael D’Rosario, Deakin Business School, Deakin University, Melbourne, Australia

## ABSTRACT

The prediction of legal outcomes and other legal domain related variables has served as the basis of a number of recent studies. While recent studies have estimated standardised variables and dichotomous outcomes such as the outcome of a judicial decision process, few studies have employed dichotomous data and categorical data to predict the basis of a legal defense strategy or the likelihood of trial success. Empirical research within the judicial sciences continues to employ a limited subset of empirical methods. This article reasserts the benefits of several artificial intelligence based non-parametric techniques that are better suited to the discipline than many of the common methods employed within the literature. The article considers the predictability of fair use defense within the U.S. during copyright infringement proceedings, and the likelihood of trial success.

## KEYWORDS

Artificial Intelligence, Copyright, Fair Use, Judicial Decisions, Law, Legal Practice, Machine Learning

## INTRODUCTION

The prediction of judicial processes has been the subject of much research in recent decades. There is a general acceptance of traditional economic methods within the sphere of legal research<sup>1</sup> (Barker, 1996). The application of economic and econometric methods in the legal domain is ever increasing. Artificial intelligence (AI) methods are generating substantial interest with the legal community. Some might argue that the interest has been a long time coming given the advent of such technologies over three decades ago. Indeed the more practical advent of new methods, such as AI technologies in recent years has made such technologies more accessible to those within the legal domain<sup>2</sup>.

The advent of such technologies has not met with positivity from all practitioners. But the potential has resulted in a heightened sense of importance and a desire to develop greater familiarity that is palpable<sup>3</sup>. This occurrence is perhaps warranting some moderation of the evidenced euphoria about such methods in legal practice<sup>4</sup>. While some contend that AI methods have the potential to replace practitioners, the present article takes a contrarian view and argues that AI technologies shall serve

DOI: 10.4018/IJSDS.2017040103

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

to supplement traditional legal information sources, and inform trial strategy. There is also scope for the emergence of a secondary market for data-driven legal services.

Professor Art Cockfield moderated a recent discussion where participant innovator and legal graduand Addison Cameron Fuff offered some insightful comments. “I think the change is people being more proactive. Right now lawyers are very reactive; somebody has an issue, and a lawyer researches it using books and databases. There is an opportunity for software to make that first pass, to highlight new issues. When a new case comes out, you shouldn’t have to wait weeks for a newsletter. It should come into your inbox. That proactive aspect is something computers can deliver, because no lawyer on the planet could possibly read all of the cases, laws and regulations that come out. Depending on your scope, you could be talking municipal, provincial, federal, international...”<sup>5</sup>

Jordan Furlong in contributing to the same recent AI and Law dialogue asserts that; “we’re going to see the adoption of AI in the legal market, more broadly speaking, rather than in the legal profession for quite some time to come. Lawyers are sort of naturally disinclined, for cultural reasons, to disrupt the way they work and go about their jobs. Technology tends to generate that aversion”<sup>6</sup>. The present article offers an example of just how effective such methods and technologies might be in support of legal practitioners.

Within the legal domain the regulatory framework and its support structure are evolving, capturing more data and enhancing courts administration and judicial accountability.

Richardson (1989) posited the advantageousness of optimisation methodology, specifically of the economic flavour when discussing the role of analysis within the courts, the author’s claims remain similarly valid to Artificial intelligence methods. Essentially courts are concerned with the allocation of resources and the behaviour of individuals. While somewhat reductionist there is truth in the claim.

It is therefore logical to assume that the rules and models of sanction should be framed while having regard for the potential incentives and disincentives these rules and pronouncements create, and their likely impact on future resource allocations (Richardson, 1989). The current article considers the extant empirical research employed within the literature positing an alternative to the common logit methods employed within legal research.

Advances of empiricism within the law and economics sub-disciplines serve as the starting point for the subsequent précis on research methods. Arguably, the most acceptance of empiricism has been observed in Intellectual property research and judicial decision-making research. This is noteworthy given the present article’s focus on both I.P. law and judicial process outcomes. For brevity’s sake, these spheres of research shall be surveyed as they represent emergent bodies of research employing empirical methodologies to legal challenges.

## **EMPIRICISM WITHIN INTELLECTUAL PROPERTY LAW**

There is limited literature on the success of fair use defenses and empirical fair use research generally (Sag, 2012). Consequently we review the broader fair use and empirical Intellectual Property literature.

Siebeck et al (1990) offers unique insight into the economic benefits of strengthened IPRs within non-developed countries. The study reviewed the extant literature both encompassing both theoretical and empirical studies. The study claims that the body of literature pertaining to industrialized economies suggests that increases in IPR protections increase R&D activities to the extent that the social cost of having IPR protections is offset.<sup>7</sup> Notably, all surveyed research employed traditional regression methodologies, none of the reviewed studies employed artificial intelligence methods.

Cotopia & Gibson (2015) In “Copyright’s Topography: An Empirical Study of Copyright Litigation” claim that one of the most important ways to measure the impact of copyright law is through empirical examination of actual copyright infringement cases (Ibid, 2015). The authors respond to the dearth through the analysis of differences between copyright and non-copyright litigation. The study employs traditional regression methods.

Menell (2016) considers the laws of fair use as they relate to mashups. The author contends that “the appropriate inquiry is whether an allocation mechanism achieves the best overall resolution of the trade-offs among authors’ rights, cumulative creativity, freedom of expression, and overall functioning of the copyright system. By adapting the long-standing cover license for the mashup genre, Congress can support a charismatic new genre while affording fairer compensation to owners of sampled works, engaging the next generations, and channeling disaffected music fans into authorized markets.” The author analyses fair use employing a conventional economics approach.

Chen (2005) considers the complex trade-off that exists between developing nations pursuing active innovation or choosing to imitate or mimic foreign rivals. The author shows that innovation within a developing country context gives rise to increases in domestic intellectual property protections. Moreover, the countries development status influences its level of IPR protection. The author employs a panel of data for 64 developing countries to “confirms both the positive impact of IPRs on innovations in developing countries and the presence of a U-shaped relationship between IPRs and economic development”(Chen, 2005).

Richardson (1991) in building upon general economic development studies asserts the worthwhile contribution associated with protection of IPRs within the framework of a closed economy; and considering the complexity deriving from the international nature of IPRs and innovation. The author provides a short exposition on the nature of the current U.S. practices and framework for IPR protection and enforcement particularly in relation to the interactions and model of sanctions employed on the basis of the national origin of an infringing service or product. The study contends that it is inapt to employ a subsidy based approach to providing incentives for R&D to promote shifting profits to U.S. firms (Richardson, 1991). The noteworthy study employs traditional economic methodology within the analysis.

McCalman (2005) titled International Diffusion and Intellectual Property Rights: An Empirical Analysis is a noteworthy study rejecting “...traditional thinking about intellectual property rights (IPR) suggests a monotonically increasing relationship between property rights and the speed of diffusion of new products and technology. Our analysis of data on the international release pattern of Hollywood movies suggests a more complex story: although moderate standards of IPR encourage the spread of movies, either weaker or stronger property rights tend to decrease the speed with which American movies are released abroad. This empirical finding is consistent with a variety of specifications, including controlling for countries’ self-selection of IPR standards. Overall, it appears that while some IPR recognition may encourage diffusion, very strong IPR may actually retard the speed of diffusion...” (McCalman, 2005). The study employs traditional regression methodology consistent with the broader literature.

Glass & Saggi (2002) explore the role of FDI and IPR’s in moderating innovation and imitation. The authors develop a model based on product life cycle with endogenous innovation while accounting for FDI and imitation. The authors “find that stronger IPR protection keeps multinationals safer from imitation, but no more so than Northern firms. Instead, the increased difficulty of imitation generates resource wasting and imitation disincentive effects that reduce both FDI and innovation. The greater resources absorbed in imitation crowd out FDI. Reduced FDI then transmits resource scarcity in the South back to the North and consequently contracts innovation.” (Glass & Saggi, 2002).

McCalman (2002) conducts an analysis of TRIP’s agreement implementation focusing on the costs, benefits, and incentives associated with the agreement. The authors determined that “the static costs of raising the standards of patent protection are captured by the transfers of income between countries, with the majority of countries estimated to make net payments abroad, the United States being a major beneficiary. To offset these transfers the model provides estimates of the dynamic benefits from the greater incentive to innovate, revealing that there is potential for all countries to benefit from the TRIPs agreement in the long run. However, the distribution of these benefits is highly skewed towards developed countries.” (McCalman, 2002).

The study of Gould & Gruben (1994) considers the theoretical relationship between IPRs and economic growth. This matter is frequently considered within econometrics research. Employing panel data, the study considers this relationship while accounting for country-specific considerations, such as trade, patent protections, and other country-specific factors, determining that rights and protections are strongly and positively associated with economic growth, which is consistent with the broader literature.

A further study by Yang and Maskus (2001) sought to respond to the shortage of research in intellectual property rights literature employing contemporary empirical methods such as panel data approaches and GMM estimation. The study sought to assess the linkage between licensing agreements and support for Intellectual Property rights IPRs. The study identifies the countries that are employing stronger patent protections and those that value patent rights generally attract a larger arm's length volume of licensed technology.

Employing traditional regression methods, Landes & Posner (2003) claimed that the extension of copyright had nominally positive but statistically insignificant effect on copyright registrations. Utilising pooled panel data the study is a noteworthy contribution, but due to the somewhat restrictive methods utilised the study cannot account for interaction effects.

The present study response to the dearth of research employing non-parametric estimation techniques in analysis the interplay between intellectual property regulation, judicial decision making, and fair use litigation.

## **EMPIRICISM WITHIN JUDICIAL DECISION MAKING**

The literature is replete with exploratory studies conducted considering the predictability of judicial decision making employing traditional linear OLS, GLS and logistic regression methods. Burton (1976) offers what is arguably the first systematic and extensive application of cross-judicial methodology. The authors examine the seminal accounts of the members of the Rehnquist Court (1986–94 terms) contrasting them with "...the prior appellate court experience to discern any correlation with their Supreme Court behaviour in terms of non-consensual opinion writing and voting..."

The authors make the remarkable discover that judges change markedly. The authors assert that the judges ... become less consensual as justices than they were as judges in the lower court. Importantly, this finding holds after controlling for such institutional differences between the two court levels as size, ideology, case types, stare decisis, and norms. Consistent with the neo-institutional perspective, we surmise that this behaviour change is due to the modern Supreme Court being unique, a court on which the members feel it is desirable, necessary, and possible to express policy disagreements with the majority via separate opinions and votes.<sup>8</sup> The study authors employ conventional regression methods in the analyses. The underlying complexity of the decision process implies that it may be non-linear in the opinion of the author.

Billica (1997) is an example of a quantitative study considering the judicial decision-making process. Specifically, Billica considered the process in the United States, by which 'states are pushed into the courts for review and interpretation.'<sup>9</sup> The study considered statutes across a 28-year period through the 86<sup>th</sup> and 99<sup>th</sup> congresses from 1959-86. Through the use of a randomly selected subset of statutes from the main dataset, the study tracks statutes that have been challenged by the courts, considering the issues addressed by the courts in interpreting statutes.<sup>10</sup>

Sisk et al (1998) considers the quandary faced by judges in 1998 when hundreds of federal district judges were suddenly confronted with the need to render a decision on the constitutionality of the Sentencing Reform Act and the newly promulgated criminal Sentencing Guidelines. Never before has a question of such importance and involving such significant issues of had constitutional law mandated the immediate and simultaneous attention of such a large segment of the federal trial bench. The authors assert that the event serves as an archetypal model for understanding the influence

of key factors such as social background, ideology, judicial role and institution, and other factors on judicial decision making<sup>11</sup>.

The authors have produced what is described as "...unprecedented empirical study of judicial reasoning in action. By exploiting this treasure trove of data, the authors have looked deeper into the judicial mind and observed the emergence of influences upon the manner in which a judge examined the constitutional issues, adopted a constitutional theory, and engaged in legal reasoning." The seminal account employs traditional regression methodology.

Yoon (2006) offers a unifying account of judicial tenure. The author claims that; Analysing judicial turnover, year by year, this article finds that judges have increasingly synchronized their departure from active service with qualifying for their judicial pension. By comparison, political and institutional factors appear to have little influence on turnover rates. These findings contradict much of the existing scholarship on judicial turnover and also offer more viable alternatives for judicial reform. The study is a further example of the style of research necessary within the legal domain<sup>12</sup>.

Martin, Quinn, and Epstein consider the matter of median justice identification.<sup>13</sup> Adapting the method framed by Martin and Quinn,<sup>14</sup> the study develops an improved quantitative method for median justice identification based on conventional regression techniques.<sup>15</sup> 'The Median Justice' is an exemplary study of interdisciplinary law and decision sciences research. Nonetheless, there remain relatively few predictive studies within law and economics, particularly in relation to judicial decision-making.

Epstein et al. consider the pertinent matter of how judicial decision making is impacted by threats to national security.<sup>16</sup> The study employs a large dataset, incorporating every civil rights and liberties case from between 1945 and 2005. The study indicates that during periods of conflict, when national security is threatened the justices are 'substantially more likely' to curtail rights.<sup>17</sup>

The study is a further example of the benefits of quantitative methods in judicial decision-making research. Like much of the research surveyed in this paper, the extant literature employs one of several popular estimation techniques. In the noted studies, logit and probit-based parametric estimation techniques are employed. The present study, as such, is a worthwhile departure from the norm, adopting a perceptron models to robustly identify associations between judicial experience and judicial outcomes and develop potentially more accurate predictive models.

Calderia and Wright consider the participation of the various parties that participate in judicial processes as *Amici curiae* within the U.S. Supreme courts, pursuant to petitions for writs of certiorari and jurisdictional statements, in addition to decisions based on merits. The authors employ a large dataset constituted by all cases involving *amici* since 1982. The study employed traditional regression methodologies.

Collins and Martinek contribute to the judicial decision-making literature by considering whether ideology mediates the impact of *amicus curiae* briefs on the decision making in the U.S. courts of appeals. The authors contend that *amicus* briefs afford the judges with critical information that influences their voting decisions. Employing conventional econometric methodology the authors assert that the *amicus* briefing influences judicial decision-making, but the influence is contingent on the judge's ideology.

The noted studies are broadly representative of the scope of works within the sphere of judicial decision-making research, from a methodological perspective, all employing traditional regression methodologies.<sup>18</sup> Herein an alternative is introduced, the multi-layer perceptron, detailing potential applications within empirical legal research.

The noted studies all employ logistic or probit modelling techniques. The most commonly used method employed in empirical judicial sciences research is logistic regression, largely because it is well suited to the dichotomous dependent variables that serve as the subject of many studies within the judicial sciences. While the contributions to empirical judicial research have been apt, many complex nonlinear problems have not been considered empirically. The present article proffers an

alternative method asserted to be parsimonious for research in the judicial sciences, representing a viable alternative to the commonly employed logit model.

## **THE NEXT WAVE OF LEGAL EMPIRICISM, ARTIFICIAL INTELLIGENCE WITHIN THE LEGAL DOMAIN**

Richardson & Hatfield (1989) describes this renewed focus on empiricism as the second wave of law and economics, with the first wave as claimed by the author evidencing only limited use of empirical methods<sup>19</sup>. While there remains the need for further acknowledgment of the benefits associated with parametric estimation methodology within the law and economics field of research amidst the so-called second wave of law and economics, it is apparent that novel non-parametric methodologies made possible by modern computing technologies represent an equally important sphere of inquiry (D’Rosario, 2016). These non-parametric methods offer insights not possible when employing traditional parametric regression. In particular, the advantages of multi-layer perceptron models in comparison to logit models when modelling complex nonlinear processes are pertinent to consider (D’Rosario, 2017).

### **A Survey of Artificial Intelligence, Specifically Perceptron Models, and Expert System Methodology Within the Legal Domain**

The Multi-layer perceptron model is asserted to be a viable supplement, and where data is limited, or relationships are complex, bi-directional and nonlinear, a superior alternative to the logit model because it offers a more accurate framework for prediction. Consider some of the challenges associated with alternative parametric techniques. Firstly, there are obvious issues associated with predictive analysis where the process is nonlinear. Logit and Probit models are relatively less capable of modelling nonlinear decision processes. MP-ANN overcomes this challenge through the application of weighted, aggregative nonlinear values based estimation. Neural networks have been advocated as an alternative modelling technique to logistic regression within the applied sciences. However there is a dearth of research considering the viability of such methods within law and economics research.

Within law and economics research, logistic regression is by far the most commonly employed method, given its suitability to dichotomous dependent variables. Within the applied sciences many more complex alternatives are being adopted. These methods present as viable alternatives to logit and probit modelling.

As asserted by Tu (1996) neural networks offer a number of advantages over logistic regression, including the ability to implicitly detect complex nonlinear relationships between dependent and independent variables, the ability to detect all possible interactions between the specified predictor variables, and the availability of multiple training algorithms. Driesietl and Ohno-Machado (2002) offer a formal methodological review of each modelling technique, noting the formal association between the techniques. The multi-layer perceptron model can be seen as an abstraction of the logistic model, employing a sigmoid function. A more apt description, however, would be that the logistic model is a simple single layer perceptron model, that does not employ a second (layer) weighted aggregate input set.<sup>20</sup> The following are studies employing artificial intelligence and expert systems methodology within the legal domain. The summation is not exhaustive but rather an outline of indicative studies worthwhile for legal scholars.

Zeleznikow (1995) contend that argument structures proposed by Toulmin can be used to represent legal knowledge in a manner that enables rule-based reasoning to be integrated with neural networks. The authors employ this approach for the construction of a system known as Split-up which predicts the outcome of property disputes in the domain of Australian family law.

Bellucci, Lodder and Zelznikow (1994) contend that current research in developing negotiation support systems focuses upon argumentation, artificial intelligence and game theory<sup>21</sup>. These techniques are rarely used in tandem. The authors argue that truly intelligent negotiation support

systems require the integration of such techniques. The authors integrate the argumentation techniques of Lodder and the combined artificial intelligence/game theory approach of Bellucci and Zeleznikow to develop an online system of negotiation.

Oatley et al. (2005) summarises research within the sphere of crime prediction outlining both the authors own work and experiences as an employee of the United Kingdom police force<sup>22</sup>. The study incorporates detailed explanatory notes from the spheres of spacial statistics and forensic psychology that serve to inform the case based reasoning and logic programing within the presented models<sup>23</sup>. The crime matching techniques used are case-based reasoning, logic programming and ontologies, and naïve Bayes augmented with spatio-temporal features (Ibid, 2005). The crime prediction techniques are survival analysis and Bayesian networks (Ibid, 2005). The techniques present favourably against traditional methods in terms of ease of interpretations and enable predictive estimation not possible with traditional methods.

Bellucci (2001) conducts expert systems research centered upon the sphere of Australian Family Law<sup>24</sup>. The authors analyse, detail and critically appraise four systems that have built in their specialized laboratory, specifically *Family\_Negotiator*, *Split\_Up*, *AdjustWinner* and *DEUS*. In doing so, we learn the complexities of the negotiation domain and investigate modelling issues for construction of an intelligent negotiation support tool.

The research has centered upon the study of negotiation strategies, specifically how to construct decision support technologies to help support human actor decision making and strategy formulation. The authors hope that their discoveries will result in future research involving multi-criteria decision making, the application of genetic algorithms and graph theory to build more efficacious Negotiation Decision Support Systems (NDSS).

D’Rosario (2017) seeks to identify whether trial representation experience possessed by lawyers within the U.S. Supreme Court impacts upon judicial decision making through the formulation of a predictive model. The author notes that a number of relatively recent studies within the extant literature have considered the role of experience in influencing judicial decision-making<sup>25</sup>. The D’Rosario study responds directly to the seminal work of McGuire<sup>26</sup>. McGuire is a significant study into the impact of the experience of competing trial lawyers on judicial decision making, employing traditional logistic regression methodologies<sup>27</sup>. The McGuire study found that trial experience possessed by trial lawyers was associated with favourable trial outcomes. The D’Rosario study extends upon McGuire, assessing the robustness of the original study employing a series of more advanced parametric estimation techniques (D’Rosario, 2017). The study then uses the McGuire logistic model framework to develop a model of prediction, employing a backward propagation, multilayer perceptron network model (D’Rosario, 2017).

D’Rosario (2016) asserts that significant methodological and theoretical contributions have been made to the extant intellectual property law and economics literature. The author asserts that ‘in recent years, a number of studies have considered the impact of IPRs on software piracy, specifically TRIPS and more recently U.S. USTR 301 reporting, pursuant to the Trade Act’ (D’Rosario, 2017). The noteworthy study of Shadlen (2005) is broadly supportive of the notion that IPR reforms directly influence rates of copyright infringement. The D’Rosario (2017) study extends upon Shadlen (2005), contrasting the traditional pooled panel model framework evidenced to a novel alternative model, a backward propagation, multilayer perceptron network model. The analysis conducted within the study focuses directly on ASEAN member countries. The study identifies the importance of IPRs and governance variables as well as presenting substantially improved predictive accuracies and goodness of fit measures.

Each of the noted studies incorporates the MP-ANN method in place of traditional estimation, to elucidate understanding in legal analysis and judicial sciences research. Pertinently each paper employs different data frameworks and variable structures. The studies incorporate binary, continuous, categorical and ordinal data, and as such introduce scholars to the treatment of each variable type. Each is written in a manner that is accessible to a non-technical audience.



## STUDY DATA, HYPOTHESES, AND METHODOLOGY

The data was procured from Sag (2012).<sup>28</sup> Sag’s seminal study compiled trial data and litigant data from the U.S. district courts, consider the main battle ground for copyright disputes, over a 33 year period between 1978 and 2011. The present study employs consolidated court determination data as derived by Sag (2012), and from the United States Supreme Court Judicial database, as presented in the Sag (2014) seminal study. A total of 222 cases ( $n=222$ ) were included in the study.

*Win\_fu* is the dependent variable denoting a successful or unsuccessful fair use defense. *Rel\_firm* denotes the relative firm sizes of each of the litigants. *Repeatplayer* is the critical repeat action variable denoting whether a defendant has appeared previously. *Useshift* accounts for a shift in the usage of the media, essentially transformative usage. *Unpublished* is a binary/Boolean variable denoting the publication status of the media. *Commuse* denotes the usage status of the media, either commercially on non-commercially. *Partial* is a variable denoting the degree of replication either full or partial. *Year* denotes the year of the trial. *While natural* denotes the natural person’s status of the defendant (see Table 1).

A number of factors are likely to have influence, aside from those considered herein, the model specifications seek to determine the critical association between litigant advantage, amicus curie and, critically, judicial experience with petitioner success. The manner with which this relationship is explored appears to be the first unique contribution of the current study, extending upon the existing work of Sag (2012).<sup>29</sup>

It is postulated herein that the relative firm size of litigants is a key determinant of judicial outcomes when accounting for litigant natural person status. A further postulate is that transformative use is positively and significantly associated with copyright fair use defense failure. The latter postulate is consistent with Sag (2012).<sup>30</sup> Consequently, we test the following hypotheses:

- H1:** Where a defendant is larger relative to the claimant in firm size terms this position shall be positively and significantly associated with litigant fair use defense.
- H2:** Where the defendant engages in transformative use this action shall be positively and significantly associated with fair use defense success.

Table 1. Variable descriptions

Variable	Variable Description	Type
<i>Win_fu</i>	The success of the defendant employing a fair use defense	Binary
<i>reform</i>	The level of advantage of the petitioner relative to the defendant (employee pool size)	Binary
<i>UseShift</i>	The transformative use status of the copyrighted work	Binary
<i>Unpublished</i>	The published status of the work	Binary
<i>Commuse</i>	The commercial status of the defendant	Binary
<i>Partial</i>	Whether the usage was partial or full	Binary
<i>Year</i>	The year of the case	Binary
<i>Natural_cr</i>	The status of the claimant as a natural person	Binary
<i>Natural</i>	The status of the defendant as a natural person	Binary
<i>Repeatplayer_cr</i>	The repeat litigation status of the claimant	3 point ordinal
<i>Repeatplayer</i>	The repeated litigation status of the defendant	3 point ordinal

The Sag (2014) study adopted a traditional logistic regression framework to build a predictive model. Herein, three predictive models are constructed using in the first instance the same methodology as employed by Sag, and then two alternative non-parametric estimation technique capable of more accurate predictive outcomes based on neural networks theory. Firstly, a multilayer perceptron network structure based on artificial neural networks theory is adopted to predict fair use defense success. Then a hybrid logit-MPANN is employed, utilising a backward propagation technique for model training. The following hypotheses pertain to the different methods of estimation:

**H3:** It is posited that the Perceptron shall achieve higher predictive accuracy than the traditional logit regression.

**H4:** The Hybrid Logit-MPANN shall achieve higher predictive accuracy than traditional logit regression and the standalone MPANN.

The MPANN approaches lend themselves well to the nature and structure of the available data. The dearth of research considering these matters is addressed by the present study which seeks to respond to omissions in the extant literature by framing a potentially more robust and accurate collection of the predictive model. A brief introduction to each of the techniques noting their benefits is detailed in the subsequent section.

This study adopts a logit model, a perceptron model and a hybrid model with the intention of understanding the impact of Repeat player status and Firm advantage on trial success and with the intent of predicting the likelihood of defence success where a fair use defence is successful, this latter task is where the multilayer perceptron model may afford the most significant benefit.

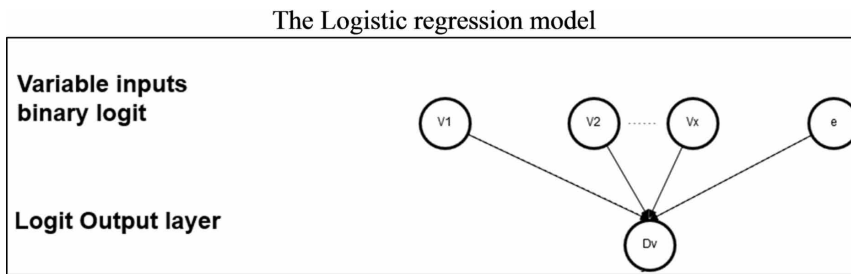
The employed data is binary, as such the excepted approach for statistical robustness all things being equal is logit regression or multiple discriminant analysis. It is posited within a number of recent studies that that such approaches are relatively less capable of accommodating dynamic and non-linear relationships. It is asserted herein that threshold effects relating to independent variables and interaction effects between independent variables cannot be estimated without knowledge of the relationship in logit specifications. A brief summary or precis on methods is provided for those unfamiliar with any of the methods, and particularly because of the new hybrid model proposed herein.

## The Logit Model

Logistic regression is a common method in empirical research, sometimes called logit modeling. The method is employed where an outcome is dichotomous in nature. When logit is employed the log odds of the dichotomous outcome is modeled as a linear combination of the predictor variables. Logistic regression seeks to measure the relationship between the dichotomous dependent variable and a vector of one or more independent variables, through the estimation of coefficients as with linear regression but by estimating probabilities using a logistic function. The logistic function has a cumulative logistic distribution.

The logistic regression model is essentially a special instance of the generalised linear model and consequently, shares a level of association with linear estimation. The underlying assumptions of the logistic regression model differ markedly from those of linear regression. The most pertinent difference relates to underlying distribution structures. The logistic regression employs a Bernoulli distribution rather than a Gaussian because the dependant variable is a dichotomous binary variable. Additionally, the predicted values of the logistic specification are probabilities and therefore seek to determine outcome values of either 0 or 1 because the output variable is binary. The logit may also be conceptually viewed as a single layer perceptron. Essentially a special instance of the perceptron without a hidden layer (see Figure 1).

Figure 1. Conceptual drawing of the logit model (Note: Here V1...Vx are the independent variables (input variables), and e denotes the error term (referred to as bias traditionally within the machine learning). Dv denotes the Dependent variable Winfu. The solid lines denote the sigmoid/logistic regression process. Dv evidences a Bi-nominal distribution).



### Model Structure

Equation (1) summarises the econometric structure of the estimation. Recall that *Win\_fu* is the dependent variable denoting a successful or unsuccessful fair use defense. *Rel\_firm* denotes the relative firm sizes of each of the litigants. *Repeat player* is the critical repeat action variable denoting whether a defendant has appeared previously. *Use shift* accounts for a shift in the usage of the media, essentially transformative usage. *Unpublished* is a binary variable denoting the publication status of the media. *Commuse* denotes the usage status of the media, either commercially or non-commercially. *Partial* is a variable denoting the degree of replication or the subject matter, copyrighted work either full or partial. *Year* denotes the year of the trial. While *natural* denotes the natural person status of the defendant, where it indexes each data item across time:

$$\begin{aligned}
 Winfu_{it} = & a + \beta.Relfirm_{it} + \beta.Partial_{it} + \beta.Transform + \beta.Commuse_{it} + \beta.Unpublished_{it} \\
 & + \beta.Repeatplayer_{it} + \beta.Natural_{it} + \beta.Repeat\_cr_{it} + \beta.Natural\_cr_{it} + \varepsilon_{it}
 \end{aligned} \quad (1)$$

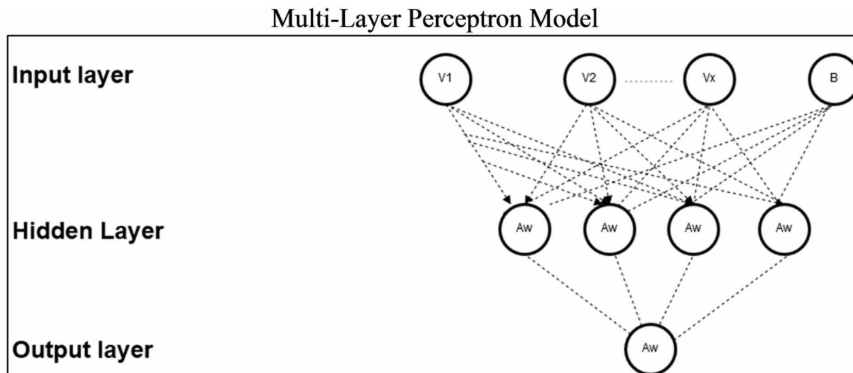
### The Multilayer Perceptron Network Model

The article given both its interdisciplinary nature and focus introduces a number of technical concepts that are likely to be unfamiliar to non-technical audiences and in particular those with a largely jurisprudential background. In the interest of engaging with the broadest audience, the following precis on neural network methodology<sup>31</sup> is provided. The multilayer perceptron model is simply a form of the artificial neural network model (herewith MP-ANN)<sup>32</sup>, a nonparametric predictive model<sup>33</sup> that seeks to replicate the structure of a biological neuron as it occurs in nature<sup>34</sup>. More plainly, it is a series of weighted, aggregative, non-linear values<sup>35</sup>. Neural network models emerged out of the body of research on artificial intelligence<sup>36</sup> that should attempt to model the human process of learning through the development of a modeling framework resembling the structure of the human brain.<sup>37</sup> The MP-ANN model has the potential to provide more accurate predictive outcomes than traditional parametric estimation techniques and is therefore well suited to the present research questions, any criticisms of legal realist logic notwithstanding<sup>38</sup>. Moreover, the methods appear more suited even accepting any legal realist underpinnings (see Figure 2).<sup>39</sup>

### Model Structure

The sigmoidal function was chosen given the data structure, and its dichotomous nature. Hidden layer size was determined based on Heaton (2011) recommendations. Optimality generally exists between the number of input layers and output layers.<sup>40</sup> Hidden layer structure was based on optimisation. Two hidden layers offered no additional optimisation benefit. Moreover, two hidden layers may be

Figure 2. Conceptual drawing of the multi-layer perceptron model (Note: Here  $V_1 \dots V_x$  are the independent variables (input variables),  $x$  denotes the number of variables and  $b$  denotes the error term (referred to as bias traditionally within the machine learning).  $Dv$  denotes the Dependent variable Winfu.  $Aw$  denote the hidden layer nodes, and the latter  $Aw$  denotes the aggregate weight at the output layer).



useful when employed in situations where modeling data evidences discontinuities. Within the present dataset, this is not the case.

The use of an additional hidden layer increased computational resources required and offered no improvement to the model. A single hidden layer model is employed to avoid a greater risk of converging to an incorrect local minima. One potential shortcoming of the selected methodology relates to the efficiency of ANN models in accommodating non-numeric variables, commonly referred to as dummy variables in economics. This is, however, only the case where non-numeric variables can take a large number of values, which is not the case in the present study.

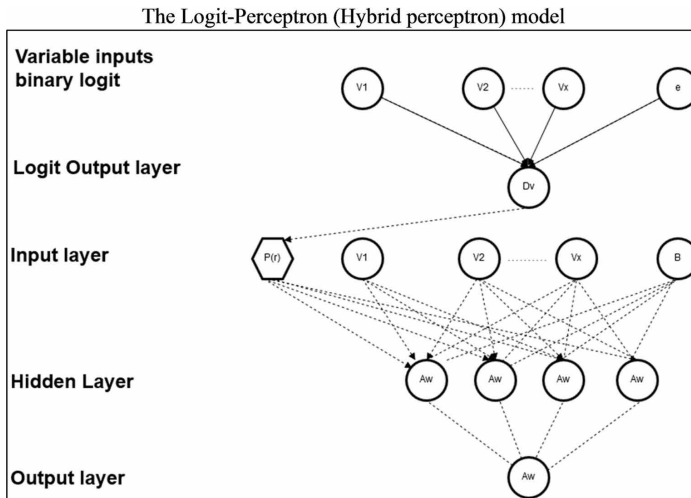
Data reduction and data pruning techniques were not necessary to improve model resolution, and the model performed efficiently and robustly absent of any node reduction. As such, the input values remain the same as those employed in Sag (2014). This is pertinent as it ensures that each predictive modeling framework can be compared on equal terms employing the same parameter set.

### The Hybrid Perceptron Network Model

The hybrid multilayer perceptron model is a new form of the artificial neural network model (herewith MP-ANN), incorporating initial logit pre-processing. The logic behind the hybrid approach is rather simple, where the perceptron has a tendency to overfitting, pre-processing may be beneficial. Herein the proposed logit-perceptron derives initial probabilities from the logit model that serve as an input for the perceptron<sup>41</sup>. Employing this method the series of weighted, aggregative values derive input from the existing vector set of variables as well as the logistic regression probability estimates. While neural network models emerged out of the body of research on artificial intelligence that should attempt to model the human process of learning through the development of a modeling framework resembling the structure of the human brain.<sup>42</sup> The MP-ANN model has the potential to provide more accurate predictive outcomes than traditional parametric estimation techniques and is therefore well suited to the present research questions<sup>43</sup>. While this article may not be the first instance where hybrid methods such as this have been employed, it appears to be quite plausibly the first within the legal domain (see Figure 3).

The MP-ANN is selected as a viable alternative to the Sag (2014) logit<sup>44</sup> model because it offers a more accurate framework for prediction. Consider some of the challenges associated with alternative parametric techniques. Firstly, there are obvious issues associated with predictive analysis where the process is non-linear. Logit and probit models are relatively less capable of modeling non-linear decision processes robustly. MP-ANN overcomes this challenge through the application

Figure 3. Conceptual drawing of the Hybrid perceptron model (Note: Here  $V_1...V_x$  are the independent variables (input variables), and  $e$  denotes the error term (referred to as bias traditionally within the machine learning).  $P(r)$  denotes the probabilities associated with the logit specification outputs.  $B$  denotes the bias function at the input layer.  $D_v$  denotes the Dependent variable Winfu.  $A_w$  denote the hidden layer nodes, and the latter  $A_w$  denotes the aggregate weight at the output layer).



of weighted, aggregative non-linear values-based estimation. The MP-ANN model appears superior when modeling complex relationships, such as the relationship between judicial experience variables and judicial outcomes.

Neural models appear more consistent with judicial decision making<sup>45</sup>. Members of the judicial system carefully weigh alternative criteria, balancing competing criteria and implicitly employ these weighted considerations in their decision making. While the process may not be explicit, it is nonetheless essential to judicial decision making. Neural networks are more capable of replicating complex decision criteria and decision weighting than the commonly employed logistic regression.

### The Layers of a Hybrid Neural Network

The first layer is the logit input layer. Here standardised and unstandardized logit inputs are employed in a logit specification. They are employed to specify a logit function representation of the relationship between the variables and litigation success. The output is a non-Gaussian distributed, Bernoulli distribution of litigation success. The logit specification classification probabilities are then captured and specified as an input in the next layer. The next layer is the first layer is the perceptron, an input layer. The perceptron input layer processes and standardises a vector of the predictor variables. This process loosely resembles the process of "saltatory conduction" with a biological neuron.

The standardised vector values of the predictor variables, reach the perceptron hidden layer. At the hidden layer, their standardised values are multiplied by a weight, with the weighted values aggregated and passed forward to the transfer function. The values from the transfer function are passed to the output layer.

The output layer accepts the values from the hidden layer, multiplying the values by weights, with the weighted values aggregated and passed forward to the transfer function. The values from the transfer function are the model outputs. As the target variable is the outcome for the petitioner, there are 2 neurons in the output layer producing 2 values, one for each of the categories of the target variable.

The hidden layer size was determined based on model optimisation, and Heaton (2011), noting optimality usually exists between the number of input layers and output layers.<sup>46</sup> Moreover, two hidden layers may be employed when modeling data that evidence discontinuities, such as a saw-tooth wave

pattern (D’Rosario, 2016). Given the nature of the underlying dataset, this is not the case. The use of an additional hidden layer reduces estimation efficiency and offers no improvement to the model in this instance. Herein, a single hidden layer model is employed to avoid a greater risk of converging to an incorrect local minima.

Data pruning techniques were not employed to improve model resolution, and the model performed efficiently and robustly absent of any node reduction. As such, the input values remain the same as those employed in Sag (2014). This is pertinent as it ensures that each predictive modeling framework can be compared on equal terms employing the same parameter set.

## MODEL FINDINGS

Considering first the logit specification, the findings indicate that relative firm size as measured by employees, the claimant’s status as a natural person and whether the defendant was a repeat player are significant at the 1% level (relsize, cl\_natural) and 10% level (df\_repeat). These findings are in contrast with the earlier Sag (2014) study that identified both the amount copied, transformative use, the plaintiff status and the defendant’s underdog status as all significant at either the five or one percent level. The contrary findings are likely due to the inclusion of the relsize variable which captures both underdog status and arguably more accurately represents firm level advantage. The substantially improved goodness of fit measures strongly supports this claim.

The multilayer perceptron findings offer some insight into independent variable importance. Remarkably, usage shifting and unpublished status are found to be relatively less important than the status of a defendant as a repeat player, the claimant status as a natural person, whether the infringing act involved partial or full copying and most pertinently the relative size variable which is asserted to be the most important determinant of fair use claim litigation success, employing the methods of Garson (1991) and Goh (1995) (see Table 2). These findings are in stark contrast with those of Sag (2012), the differences are likely due to the inclusion of the relsize variable and the method

Table 2. Summary of logit model

	B	S.E.	Wald	df	Sig.	Exp(B)
case_id	0.010	0.016	0.389	1	0.533	1.010
year	-0.128	0.194	0.437	1	0.509	0.879
cr_repeat			1.346	2	0.510	
cr_repeat(1)	-1.056	0.990	1.137	1	0.286	0.348
cr_repeat(2)	0.079	0.446	0.031	1	0.859	1.082
df_repeat			3.428	2	0.180	
df_repeat(1)	-0.284	1.041	0.074	1	0.785	0.753
df_repeat(2)	-0.898	0.492	3.325	1	0.068	0.407
useshift2	0.385	0.459	0.703	1	0.402	1.470
unpublished	0.640	0.609	1.101	1	0.294	1.896
relsize	-0.009	0.003	8.859	1	0.003	0.991
cr_natural(1)	-1.783	0.469	14.464	1	0.000	0.168
df_natural(1)	-0.867	0.601	2.083	1	0.149	0.420
Constant	256.471	385.212	0.443	1	0.506	2.42+111

Note – This table summarises the co-efficient, t statistics, and associated probabilities for all variables within the specification

of estimation. The perceptron is capable of modeling complex non-linear processes and capturing variable interaction effects (Garson 1991, Goh 1995).

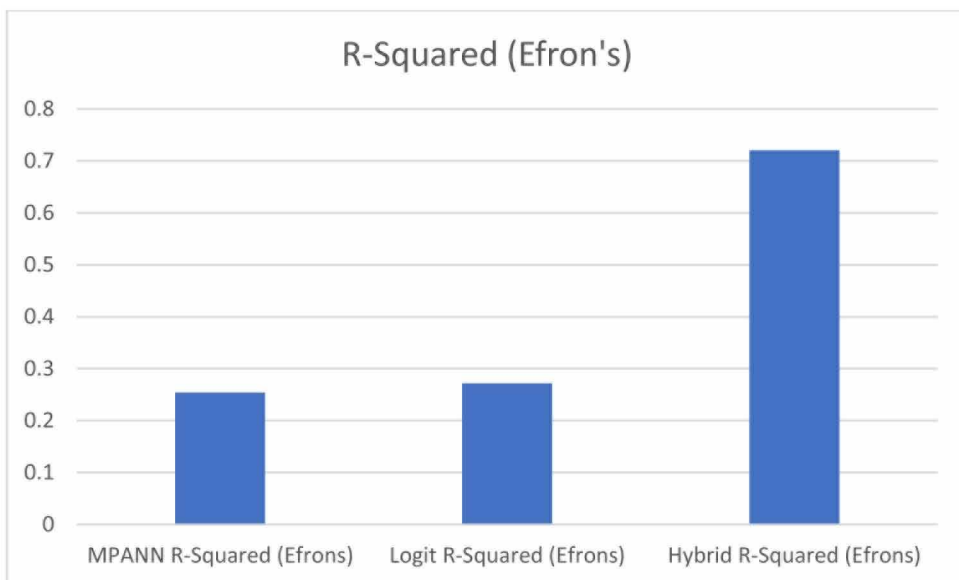
The hybrid model findings are largely consistent with the multi-layer perceptron findings, however the logit probabilities largely ranked first across all iterations of the hybrid model indicating that logit pre-processing was indeed worthwhile the relsize variable and the claimant and defendant status as repeat players were also found to be important employing the methods of Garson (1991) and Goh (1995).

### Goodness of Fit and Predictive Accuracy

We employ Efron's method of determining goodness of fit because of its ease of interpretation (see Figure 4). Unlike other measures of goodness of fit Efron's method enables the resultant R-squared to be interpreted as a measure of explained variability within the specification. The ratio is the proportion of variability explained by the model specification. It is also mathematically the square of the correlation between the actual values and those predicted by the model. Consequently, this measure observes a range from minus one to plus one. This is consistent with the equivalent non-pseudo R-squared values employed in OLS regression.

The initial multi-layer perceptron model evidence an Efron's R-squared of 0.2401. This is substantially higher than the R-squared values observed within the extant literature. In Sag (2012), R-squared values vary between 0.095 and 0.215. These logit specifications were replicated to ensure consistency when framing the alternative logit, MPANN and hybrid logit/MPANN. This represents an 11.67% increase in goodness of fit between the full form logit model in Sag (2012) and the equivalent MPANN. Considering again the models presented above, we observe a shift in predictive accuracy and goodness of fit through the inclusion of a re-specified variable relsize denoting relative firm size. The inclusion of relsize, in the logit model gives rise to substantial improvements in the logit model goodness of fit measures. Efron's R-squared, increases from 0.215 to 0.2714 a 26.25%. However, the increase in goodness of fit is not as substantial within the MPANN specification, increasing from 0.2401 to 0.2540.

Figure 4. Efron's R-squared goodness of fit measures



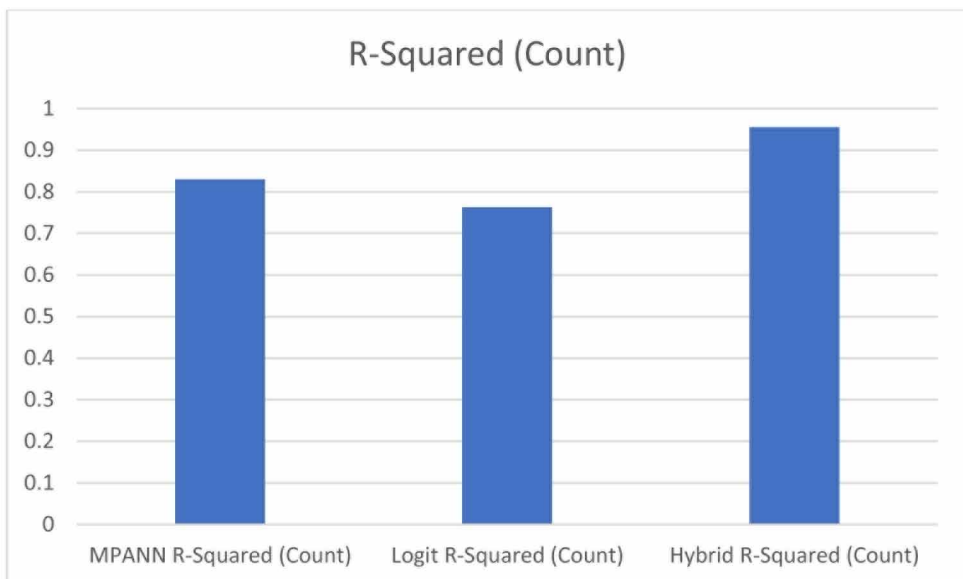
The application of the hybrid method results in some remarkable efficiency gains in terms of computational time as well as goodness of fit and predictive accuracy. Through the use of the logit model as a pre-processing method, the MPANN incorporating the logit probabilities evidences an Efron's R-squared of 0.7208 representing a 201% increase in goodness of fit, when compared with the MPANN in isolation, and 184% increase in goodness of fit when compared with the logit used in isolation. While there is no perfect measure to assess model goodness of fit, all of the indicia suggests the hybrid method (logit-MPANN) offers the most robust specification. Considering now the central question of predictive accuracy, the findings indicate that the multi-layer perceptron and the hybrid method results in far greater predictive accuracy. Consider the count R-squared values for each of the models, essentially the percentage of observations predicted correctly by each model specification (see Figure 5). The traditional logit gives rise to a count R-squared of 0.763, while the multi-layer perceptron evidences a count R-squared of 0.83 equating to approximately 9.1% increase in predictive accuracy through the use of multi-layer perceptron. Through the application of the logit model as a pre-processing technique, the hybrid multi-layer perceptron achieves a predictive accuracy level of 0.9553, a substantial increase in predictive accuracy of 15.1 to 25.2% when compared with the logit and multi-layer perceptron respectively.

Considering the posited hypotheses, the findings are broadly consistent with the stated hypotheses. Each of the posited hypotheses is noted below:

**H1:** Where a defendant is larger relative to the claimant in firm size terms this position shall be positively and significantly associated with litigant fair use defense.

The findings across all specifications suggest that relative firm size is one of the most significant determinants of litigation success. Within the logit, perceptron and hybrid-perceptron models, the coefficient and independent variable importance analysis indicate that relsize ranks first in importance terms and is the most statistically significant within the logit specification.

Figure 5.Count R-squared values for each model specification





**H2:** Where the defendant engages in transformative use this action shall be positively and significantly associated with fair use defense success.

In relation to transformative use the findings are mixed. Useshift is found to be nominally important within the perceptron models and positively associated with fair use defense success during litigation. However, it is found to be statistically insignificant.

**H3:** It is posited that the Perceptron shall achieve higher predictive accuracy than the traditional logit regression.

**H4:** The Hybrid Logit-MPANN shall achieve higher predictive accuracy than traditional logit regression and the standalone MPANN

Hypotheses 3 and 4 shall be considered concurrently. The perceptron and hybrid-perceptron achieved vastly superior predictive accuracy than the logit model. The hybrid-perceptron achieved the highest predicative accuracy overall, consistent with the stated hypotheses.

As such, in relation to the central consideration, the framing of a predictive model of fair use litigation success, the findings strongly favour the use of non-traditional methods specifically the multilayer perceptron and the hybrid multilayer perceptron. This is consistent with a number of studies incorporating discussion and/or analysis of logistic regression and machine learning methods (Tu, 1996; D’Rosario, 2016). There is however, a genuine dearth of research incorporating an analysis of hybrid methods and traditional methods within the legal domain. The present study appears to be the first within the area of litigation analysis.

## **CONCLUSION, LIMITATIONS, AND SCOPE FOR FUTURE RESEARCH**

A number of studies have evidenced the benefits of artificial intelligence and machine learning methods within the legal domain. The current study extends upon this literature offering a unifying account, and detailing a new method of estimation not utilised within the legal domain previously. Through the use of a new specified variable capturing relative firm size it was identified that the status of the claimant as a natural person, defendant repeated status and relative firm size are identified as of significance to litigation success. The perceptron and logit-perceptron offered deeper insights. The logit-perceptron is shown to achieve better predictive accuracy and better goodness of fit outcomes than the conventional regression methods employed within the extant literature.

Moreover, through the use of the perceptron and logit perceptron the variable relsize was found to be the most significant determinant of litigation success. The defendant status as repeat players was also found to be important employing the methods of Garson (1991) and Goh (1995). The findings appear to validate and further endorse the use of these contemporary non-parametric methods. The study does have some limitations worth noting specifically the limited available data and the absence of a singularly accepted method for determining robustness. Clearly, as more data becomes available, subsequent studies shall respond to the former issue in relation to the latter the use of Efron’s R-Squared and the count R-Squared method appear to be the most valid and allow for the most coherent comparison of the parametric and non-parametric methods.

While economic methods are employed consistently within legal research, the use of non-parametric methods such as those within the domains of artificial intelligence and machine learning are relatively under utilised within the legal domain. The present study seeks to advise the next wave of interdisciplinary legal research, typified by applications of machine learning within the legal domain. We refer to this next wave as the third wave of legal empiricism in honour of Richardson & Hatfield (1989). We offer the conjecture that legal empiricism shall be strongly influenced by machine learning methodology, particularly as such technologies become more accessible to those

within the legal domain. As this study makes evident, there is a dearth of research considering the relationship between law and machine learning. While this study responds directly to this shortage, further research is necessary. It would be worthwhile to consider the importance of judiciary member characteristics on judicial decision making as well as considering the legal domain beyond intellectual property. Such studies would serve to inform legal practice regulatory policy and our understanding of judicial decision making. Such studies would serve to inform the development of even more accurate and robust machine learning methods; and it is hoped that such studies are inspired and informed by the present study.

## REFERENCES

- Baker, L.A. (1996). Interdisciplinary Due Diligence: The Case for Common Sense in the Search for the Swing Justice. *Southern California Law Review*, 70, 187.
- Epstein, L., Ho, D. E., King, G., & Segal, J. A. (2005). The Supreme Court During Crisis: How War Affects Only Nonwar Cases. *New York University Law Review*, 80.
- Garson, G.D. (1991). Interpreting Neural-Network Connection Weights.
- Goh, A.T.C. (1995). Back-Propagation Neural Networks for Modeling Complex Systems. *Artificial Intelligence in Engineering*, 9(3), 143-151.
- Gressman, E., & Stern, R. (2002). *Supreme Court Practice: For Practice in the Supreme Court of the United States* (8th ed.). Bureau of National Affairs.
- Hunter, D. (1994). Looking for law in all the wrong places: Legal theory and legal neural networks. In *The relation with legal theory* (pp. 55-64).
- Hunter, D. (1999). Out of their minds: Legal theory in neural networks. *Artificial Intelligence and Law*, 7(2-3), 129-151.
- Enns, K. (1998). Can a California Litigant Prevail in an Action for Legal Malpractice Based on an Attorney's Oral Argument before the United States Supreme Court? *Duke Law Journal*, 48(1), 111-146.
- Karnin, E.D. (1990). A Simple Procedure for Pruning Back-Propagation Trained Neural Networks. *IEEE Transactions on Neural Networks*, 1(2), 239-242.
- Martin, A.D., Quinn, K.M., & Epstein, L. (2005). The Median Justice on the US Supreme Court. *North Carolina L. Rev.*, 83, 1275.
- Masko, K.E. (2002, August 6) Intellectual Property Issues for the United States and Japan: Disputes and Common Interests. In *Proceedings of the Issues and Options for the Multilateral, Regional, and Bilateral Trade Policies of the United States and Japan Conference*.
- McAtee, A., & McGuire, K. T. (2007). Lawyers, Justices, and Issue Salience: When and How Do Legal Arguments Affect the US Supreme Court?. *Law & Society Review*, 41(2), 259-278.
- McGuire, K. T. (1995). Repeat players in the Supreme Court: The role of experienced lawyers in litigation success. *The Journal of Politics*, 57(1), 187-196.
- Minsky, M., & Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. Cambridge, Mass.: MIT Press.
- Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological modelling*, 154(1), 135-150.
- Patterson, D., & Patterson, A. N. N. (1998). *Theory and Applications*. Singapore: Prentice Hall.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1988). Learning Internal Representations by Error Propagation. In A. Collins & E.E. Smith (Eds.), *Readings in Cognitive Science* (p. 399). San Mateo, CA: Morgan Kaufmann Publishers.
- Sankar, A., & Mammone, R. J. (1991, July). Optimal pruning of neural tree networks for improved generalization. In *Proceedings of the Seattle International Joint Conference on Neural Networks IJCNN '91* (Vol. 2, pp. 219-224). IEEE.
- Sheehan, R. S., Mishler, W., & Songer, D. R. (1992). Ideology, status, and the differential success of direct parties before the Supreme Court. *American Political Science Review*, 86(2), 464-471.

Songer, D. R., & Sheehan, R. S. (1992). Who wins on appeal? Upperdogs and underdogs in the United States courts of appeals. *American Journal of Political Science*, 36(1), 235-258.

Epstein, L., & Knight, J. (1997). *The Choices Justices Make*. Congressional Quarterly Inc.

Weigend, A. S., Rumelhart, D. E., & Huberman, B. A. (1990). Back-propagation, weight-elimination and time series prediction. In *Proceedings of the 1990 Connectionist Models Summer School* (Vol. 105). Morgan Kaufmann.

## ENDNOTES

- 1 See Barker (1996)
- 2 See the recent CNBC article titled “Lawyers could be replaced by artificial intelligence”; <http://www.cnbc.com/2017/02/17/lawyers-could-be-replaced-by-artificial-intelligence.html>
- 3 Ibid
- 4 Ibid
- 5 For the full plenary session transcript visit; <http://law.queensu.ca/how-will-artificial-intelligence-affect-legal-profession-next-decade>
- 6 Ibid
- 7 W E Siebeck et al, ‘Strengthening Protection of Intellectual Property in Developing Countries: A Survey of the Literature’ (1990), World Bank discussion paper No.112, Washington D.C. The World Bank.
- 8 Burton M. Atkins, and Justin J. Green . 1976. “Consensus on the United States Courts of Appeals: Illusion or Reality?” *American Journal of Political Science* 20(11):735–48.
- 9 Nancy Ellen Billica, ‘Just Leave it to the Courts: How, When, and Why Congress Abdicates the Legislative Power’ (PhD Thesis, Harvard University, 1997).
- 10 Ibid.
- 11 Sisk, Gregory C. and Heise, Michael and Morriss, Andrew P., Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning (November 1, 1998). *New York University Law Review*, Vol. 73, No. 5, 1998. Available at SSRN: <https://ssrn.com/abstract=2447727>
- 12 Albert Yoon Pensions, Politics, and Judicial Tenure: An Empirical Study of Federal Judges, 1869–2002, *Am Law Econ Rev* (2006) 8 (1): 143-180
- 13 Martin, Quinn, & Epstein (2005).
- 14 Martin, A.D., & Quinn, K.M. (2002). Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the US Supreme Court, 1953–1999. *Political Analysis*, 10(2), 134-153.
- 15 Martin et al, above n 11.
- 16 Epstein, Ho, King et al., 2005
- 17 Ibid.
- 18 Clearly there are numerous regression based methods employed within empirical judicial sciences research, and empirical legal research generally. The author does not assert that all studies employ the noted methods but rather the noted methods are broadly representative of the most commonly employed methods.
- 19 Megan Richardson and Gillian Hadfield, *The second wave of law and economics*, Federation press, 1999
- 20 For a full methodological review, see Dreiseitl, S. & Ohno-Machado, L. (2002, October). Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. *Journal of Biomedical Informatics*, 35(5–6), 352-359.
- 21 Bellucci, Lodder and Zelznikow (1994)
- 22 Oakley et. al. (1995)
- 23 Ibid (1995)
- 24 Bellucci (2001)
- 25 D’Rosario (2016)
- 26 D’Rosario (2016)
- 27 D’Rosario (2017)
- 28 The author is grateful for Matthew Sag’s generosity in making this dataset available, and thus making the present study possible.
- 29 Sag (2012), above n 2.
- 30 Sag (2012), above n 2.
- 31 A multilayer perceptron model is selected as an alternative to the radial basis function due to its superiority in estimation. Any potential gains in estimation efficiency associated with the radial basis method are offset by the losses in estimation accuracy.

- 32 McCulloch, W.S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- 33 D’Rosario (2016) “Rosenblatt posited the perceptron model, framing the concept of the retina layer with distributing input values. The work of Rosenblatt owed much to the initial work of Hebb, and also McCulloch, and Pitts. The work of Marvin Minsky and Seymour Papert asserted a number of the limitations of the perceptron model, and their work brought about a diminution of interest in perceptron modelling. The key challenge evidenced in the perceptron model as posited by Rosenblatt was the step function approach that made neural network models problematic to train. However interest in the field was renewed with the advent of techniques to moderate the issues associated with neural network training. A revival of sorts occurred as a result of the work of Rumelhart, Hinton and Williams, whose seminal paper ‘Learning Internal Representations by Error Propagation’ offered resolution. Their work enabled the effective training for a multilayer neural network. The backpropagation training method and a viable algorithm was first asserted in Rumelhart, Hinton and Williams, and was the first practical method for training neural networks. Herein, a gradient descent backpropagation training method is used for efficiency.”
- 34 Within a biological neuron, the synaptic signals are received by the soma (cell body) and dendrites and transmitted through the axon. Within an artificial neuron the vector of the variable is passed from one, neuron in the input layer to another neuron with the assignment of synaptic weights (weighted values). Transfer occurs through the assistance of a transfer function, the synthetic axon.
- 35 Minsky, M., & Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. MIT Press.
- 36 Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart, & J.L. McClelland (Eds), *Parallel distributed processing: explorations in the microstructure of cognition* (Vol. 1, pp. 318-362). MIT Press
- 37 Patterson, D.W. (1996). *Artificial neural networks: theory and applications*. Prentice Hall.
- 38 It is necessary to discuss the criticisms of neural network methodology within legal research of which there appear to be few of a unique nature. Dan Hunter details the problems of using statistical sub-symbolic techniques in a symbolic field of research (D’Rosario, 2017a). At the outset it is apparent that such criticisms are relevant to all quantitative methods not merely machine learning. Hunter, is highly critical legal realist philosophy insofar as machine learning methods such as perceptron models and artificial intelligence methods derive from them (D’Rosario.). It is evident that Hunter’s assertions are based on what Hunter considers to be inapt realist claims about the focus of network applications in his view and not the methods directly (D’Rosario, 2017). Hunter rightly and notably claims that useful applications are likely to be observed in latter studies (Hunter, 1994). It is hoped that the present study is such a study. It is also worth acknowledging that at the time that Hunter’s paper article was authored, computing technologies did not enable the use of complex multi-layer perceptron models, nor was it possible to account for independent variable importance and as such the methods available now are far more useful in accommodating symbolic and sub symbolic languages as well as the open texture within legal research. Herein, as with D’Rosario (2016, 2017a) the author agrees with the claims of Hunter only to the extent that legal realists espouse that neural networks can provide normative assessments of legal outcomes and jurisprudence. No such claim is made herein. However, Neural networks methodology does enable a fuller analysis of judicial outcomes, jurisprudence and the importance of different variables in relation to such determinations.
- 39 Consider the statements of Warner (1989), plausibly the most ardent proponent of this position, “While our language dictates a sequential description of the (legal reasoning) process, the process is in fact parallel. Many aspects of the problem resolution process are carried out simultaneously. The problem domain is defined by the initial statement of the problem. That initial problem is then resolved into a number of issues... the solution to which will be sought within the problem domain utilizing a sub symbolic paradigm that is not rule based...”
- When accounting for even this staunch position, or when ascribing to the Hunter (1994) view position, the view of Warner does not coincide with the correct interpretation of legal reasoning as a process, neural networks nonetheless remain superior in methodological terms within legal research pertaining to legal reasoning; when compared with traditional regression methodology. This is because traditional regression models such as logit models are incapable of modelling dynamic nonlinear functions and as previously noted accounting for the interaction between variables (D’Rosario, 2016, 2017).
- 40 Heaton, J. (2011). *Introduction to the Math of Neural Networks*. Heaton Research, Inc.
- 41 Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386- 408 .
- 42 Patterson, D.W. (1996). *Artificial neural networks: theory and applications*. Prentice Hall.
- 43 For the initial theoretical propositions that served as the basis for the methodology see Hebb, D.O. (1952). *The organisation of behaviour: a neuropsychological theory*. Wiley.

- <sup>44</sup> The multi-layer perceptron model shares a level of association with the logistic model, employing a logistic/sigmoid function. A more apt description however would be that the logistic model is a simple single layer perceptron model, that does not employ a second (layer) weighted aggregate input set.
- <sup>45</sup> See Lothar & Philipps (1981), Hobson & Slee (1994) and Rose & Belew (1989)
- <sup>46</sup> Heaton, J. (2011). *Introduction to the Math of Neural Networks*. Heaton Research, Inc.

*Michael D’Rosario is a lecturer and unit chair in post graduate programs at Deakin business school. Prior to working at Deakin business school Michael worked with Price Waterhouse Coopers, Ford Motor Company, Korda Mentha, and the University of Melbourne. Previously, Michael served as a technical advisor to data science related projects and presently serves as deputy chairman of Stirling Theological College, Melbourne University of Divinity. Michael is an interdisciplinary researcher having published in the domains of law, econometrics, decision sciences, theoretical economics, and information systems.*