

ARTICLE

Jurisprudence and the Intelligible World: Exploring Predictive Modelling as a Mechanism to Decide Bail in the Australian Context

Brett Anthony Hansard^{id} and Jianlong Zhou^{id}

School of Computer Science, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, New South Wales, Australia

Corresponding author: Brett Anthony Hansard; Email: brett.a.hansard@student.uts.edu.au

(Submitted 2 June 2025; revised 5 June 2025; accepted 8 June 2025)

Abstract

The intelligible world of machines and predictive modelling is an omnipresent and almost inescapable phenomenon. It is an evolution where human intelligence is being supported, supplemented or superseded by artificial intelligence (AI). Decisions once made by humans are now made by machines, learning at a faster and more accurate rate through algorithmic calculations. Jurisprudent academia has undertaken to argue the proposition of AI and its role as a decision-making mechanism in Australian criminal jurisdictions. This paper explores this proposition through predictive modelling of 101 bail decisions made in three criminal courts in the State of New South Wales (NSW), Australia. Indicatively, the models' statistical performance and accuracy, based on nine predictor variables, proved effective. The more accurate logistic regression model achieved 78% accuracy and a performance value of 0.845 (area under the curve; AUC), while the classifier model achieved 72.5% accuracy and a performance value of 0.702 (AUC). These results provide the groundwork for AI-generated bail decisions being piloted in the NSW jurisdiction and possibly others within Australia.

Keywords: bail; decision-making; criminal courts; machine learning; artificial intelligence; binary logistic regression; tree-structured classifier

Introduction

Contained within the Bail Act 2013 (NSW) is a dichotomous or binary decision-making process as to whether an individual accused of one or more offences is granted or refused bail – a decision entrusted to a bail authority, defined within the Bail Act 2013 (NSW) as an authorized justice of a court – and that an “accused person” (also known as a “defendant”) would need to demonstrate to that bail authority the reasons why their detention is unjustified (Bail Act 2013 (NSW)). Not long after the Bail Act

2013 (NSW) came into force, what appeared to be a retrospective measure to strengthen it was the Bail Amendment Bill 2014 (NSW), the Second Reading of which by the then Attorney-General stating that an accused charged with a specified level of offending seriousness would need to “show cause . . . that detention is not justified and be subject to the unacceptable risk test before bail can be granted” and that a bail decision (used interchangeably with bail judgment) would be made based on an “unacceptable risk assessment” (Bail Amendment Bill 2014 (NSW)). The Attorney-General explained further that the matter of “bail concern” is indicative of five items: failing to appear; committing a serious offence; endangering the safety of victims; the community; or interfering with witnesses of evidence (Bail Amendment Bill 2014 (NSW)).

Decision-making is a fundamental undertaking of criminal courts. It is the ideal expectation that corresponding cases should result in consistent decisions, as remarked upon in the Second Reading of the Bail Amendment Bill 2015 (NSW). Despite being legislated and systematized, it is more realistic to expect inconsistencies in decisions to occur. What if the courts themselves consider consistency, balance and efficacy when making bail decisions? The following extracts have been taken from judgments on bail cases that lead to this question:

It is possible, if not on one view, that minds may differ in any particular case about how these authorities should play out in the difficult discretionary exercise with which I am presently concerned (*R v. Peter Tsallas* 2017, 978).

To further demonstrate how inconsistencies are a factor in judicial decisions, a similar position was expressed in the New South Wales (NSW) criminal appellate court matter:

Bail decisions involve a discretionary evaluative judgment on a variety of factors about which, and within limits, reasonable minds may differ. However, every bail application presents its unique factual matrix (*DPP (NSW) v. Zaiter* 2016, 247).

In adversarial criminal justice systems such as NSW, bail decisions are not required by law to meet the standard of proof that convictions require, namely, “beyond a reasonable doubt”. As indicated in the Bail Act 2013 (NSW) in Section 31, “Rules of evidence do not apply” – a bail authority is only required to meet the standard of proof on a “balance of probabilities” – accordingly, there is not any requirement to meet the reasonable doubt standard when making a decision, other than determining new and untested criminal matters (Bail Act 2013 (NSW)). While the literature consistently calls for high accuracy in outcomes, a benefit presumably gained from Section 31 of the Bail Act 2013 (NSW) in probabilistic terms is that decisions would not have to meet the principle of absolutism, as perhaps might be expected in medical assessments; rather, the principle of probabilism is more applicable. A total of 24 legislative provisions guide bail authorities in determining bail. The NSW Supreme Court expressed the relevance and weight of these 24 factors in the following extract:

... [section] 18 limits the matters to be taken into account in assessing bail concerns under the [Bail] Act. Each matter is given equal priority. No one matter assumes dominant significance (*JM v. R* 2015, 978).

In a bail decision handed down in the NSW Supreme Court, the judge provided a succinct characterization of how bail is decided under the new legislation and subsequent amendments and referenced the statement made by the Attorney-General, described earlier:

[T]he approach of the Court falls into a dichotomy. If there is an unacceptable risk, the Court must refuse bail; if there is no unacceptable risk, the Court must grant bail. That test applies to all offenses (*JM v. R* 2015, 978).

The outcome sought from this human predictive assessment is to decide if bail is granted or refused – a dichotomous decision – and therefore logical for the most suitable measure when applying machine-driven predictive modelling also to be dichotomous or binary. As such, the favoured statistical methods are binary logistic regression (B-LogR) and the tree-structured classifier (T-sC), both of which will be discussed in greater detail later. The rationale for applying two statistical methods is based on the principle of acceptable practice, where multiple methods are employed to determine the most suitable approach (McCue 2014).

The literature on AI in criminal justice has gained significant momentum globally over the past two decades, with a notable trend in Australia. The momentum has also carried over to state governments, which have established policies and frameworks on AI, notably in NSW. However, it appears that governments, including the NSW government, have yet to capitalize on this momentum and further test AI-generated decision-making in the criminal justice domain, such as when determining bail. A benefit of encouraging AI-generated bail decisions, for instance, is that it can lead to arguments based on improving budgetary measures and mitigating human error. Of discernible interest and importance to note in any considerations on AI being used in decision-making, particularly on mitigation to human error, is the ethicality of AI, for example, fairness and biases. In the current study, consideration is given to all of the aforementioned aspects, with particular interest in bail in the NSW criminal jurisdiction, in support of piloting AI-generated decisions to determine bail.

This paper's sectional arrangement commences with the introduction, from where the background and related work from the literature will be detailed. Next, the statistical methods will be outlined, including the data and classification metrics used in the study. Results from the study are then detailed, its outcomes culminating in a discussion and perspectives on future work.

Background and Related Work

A report on Australian prisons highlighted several significant issues related to imprisonment rates and court administration processes (Productivity Commission 2021). For example, the rate of remandees awaiting court outcomes

is almost twice the number it was two decades later than in the year 2000, and the average time in custody has increased by 1.3% to 5.8 months over the same period, from 2000 to 2020 (Productivity Commission 2021). More specifically, in NSW, the average remand time increased from previous years to 6.1 months in 2020; one factor blamed for this increase was systematic issues with court processing (Productivity Commission 2021).

Following the initial bail amendments in 2014, defendant numbers fluctuated in all adult-based court levels in NSW. Correspondingly, court delays – which can otherwise be understood as defendants remanded – increased in the Local and District Courts, and the Supreme Court showed some sharp variations (NSW Bureau of Crime Statistics and Research 2023). Essentially, an implication from these data is that defendants awaiting court outcomes are doing so while incarcerated, at the risk of adverse administrative and legal consequences, only to find themselves released without any further penalization.

During the period 2018 to 2020, defendants in NSW adult criminal courts whose status was “bail refused” in all three court levels of local, district and supreme courts are equally represented in most categories. Although two categories of “not guilty of all charges” at 5.57% ($n = 579$) and “all charges withdrawn by the prosecution” at 4.74% ($n = 493$), while proportionate to the number ($n = 10,393$) of matters finalized (NSW Bureau of Crime Statistics and Research 2023), may not seem overwhelming, the implication is that defendants were remanded without any form of sentence being handed down by the courts despite lengthy periods of incarceration. An inference that may be drawn from these two categories is that bail authorities, such as magistrates and judges, were incorrect in remanding some individuals rather than granting conditional liberty. A further discussion could be had on other categories where defendants were remanded and whether bail would have been a more suitable option, for example, than incarcerating mentally ill or cognitively impaired persons, although that topic is not of direct relevance here.

Apart from the legal and moral issues associated with erroneous decisions, another concern is the cost-effectiveness of potentially unnecessary incarcerations. In NSW, the budget concerning incarceration is increasing (Audit Office of New South Wales 2019), and over the 2014–2015 financial year – not long after the Bail Act 2013 (NSW) and subsequent amendments came into effect – the net cost to keep an individual in custody in Australia was calculated to be \$61,179 annually as opposed to approximately \$6,500 annually in the community under court-issued orders (Morgan 2018). The assumptions made from these figures are based on cost-benefit efficiencies: an accused on bail would cost significantly less than one on remand and, again, much less than community-based orders, as generally, there is limited monitoring and intervention required for individuals on bail. Notwithstanding the cost-reduction argument, another issue is the time on remand, which was not solely risk-based but also attributed to laborious court processes and erroneous decisions.

Considering the economic benefits and efficiencies of sentencing, jurisprudent scholars have argued in favour of a progressive measure to alleviate this issue, namely, the use of AI. Stobbs, Hunter, and Bagaric (2017) commented that algorithms designed to undertake such a task are fiscally responsible and would create efficiencies in court administration in addition to other benefits of

consistency, transparency and predictability. Subsequently, the objective is to improve the efficiency and accuracy of the criminal justice system. Bail decisions, conventionally made by the judiciary, could be replaced with contemporary methods, inferring that AI could substitute human intelligence.

Conventional to Contemporary Predictive Instruments to Decide Bail

A former chief justice of the High Court of Australia, in a speech on the current and future implementations of technology in Australian courts, opined:

Technology is an integral part of our daily lives. It is the now and the future. One does not need to look too far to see mistaken disregard for technology in the past (Allsop 2019).

More specifically, AI, as a prominent technological influence, has been declared on a macro or global scale, bringing about transformations in economies and workplaces through increased productivity and innovation and potentially enriching the lives of people and societies (Organisation for Economic Co-operation and Development 2025). It is essential, nonetheless, to draw from the literature a meaning of AI, expressed in the guidelines by the Organisation for Economic Co-operation and Development (2025:7):

An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

A definition cited in a criminal justice aspect was that AI is “the scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines” (Atkinson in Dupont et al. 2018:9). Meaningfully, the objectives are human-driven, yet programmed into a machine to make predictions that have real-world implications, and therefore, the technological evolution is undeniable, and AI is going to be integral.

AI-Generated Decision-Making Prototypes

Earlier work on a prototype for decision-making in the criminal justice jurisdiction was conceptualized as a decision support system for sentencing in the Magistrates’ Court of Victoria (Farmer, Parsons, and Bagaric 2018). The authors’ decision-support system was said to apply a statistical algorithm approach, where data from previous sentencing outcomes or matching similar cases would inform decision-makers. The authors stated that the algorithm framework integrated decision trees and argument trees. Once satisfied with the theoretical model, the authors implemented it in a practical model using a web-based program, which generated responses or prompts based on weighted variables to determine the sentence ultimately. Another literature source suggests that the need for consistency and

efficiency underpins the motivation for a decision-support machine. Their argument for consistency was based on the probability that decision-makers examining the same facts can make different decisions (Hall et al. 2005). The limitations expressed by the authors regarding their prototype were more legal than technical, which was likely because their research had not tested the technical elements on actual cases; consequently, its effectiveness could not be assessed (Hall et al. 2005). Notwithstanding, the authors demonstrated that a schematic model can become an algorithm prototype.

A survey from the literature considered predicting the outcomes of criminal cases using various ML classifiers. Shaikh, Sahu, and Anand (2020) referred to their research undertaking as a “legal approach”, which surveys the narratives of relevant legal cases in search of specific text or phrasing that creates the features. The relevant information is entered into a database upon which the chosen ML classifier analyses and provides the outcome. The authors selected eight ML classifiers to analyse and predict the outcomes of 86 cases (for the full list, see Shaikh et al. 2020). It was determined that the eight classifiers correctly predicted 64 cases, yet using classification and regression trees (CART) was the most accurate (91.86%) and had the best performance (91.76%); the authors reported that all eight classifiers provided respectable predictions in accuracy (between 85% and 92%). Also, their research showed 22 cases as having a minimum of one incorrect prediction and two cases wrongly predicted by all the eight classifiers. A limitation of the research by Shaikh et al. (2020) was identified in the high number of arbitrary predictors and descriptors. As contended elsewhere, weak predictors can be counterintuitive (Berk and Bleich 2013).

Other scholars have explored a similar research question to the one previously mentioned, although a discernible difference was apparent regarding the most effective model. Zeng, Ustun, and Rudin (2017) hypothesized that the ML classifier performance of the Supersparse Linear Integer Model (SLIM) would provide superior accuracy and interpretability over eight other classification models for predicting recidivism, including CART (for the eight models, see Zeng et al. 2017). These authors thought that SLIM is well-suited for quantitative research in criminology; it has similarities to conventional linear risk assessments for recidivism where a user can observe what input variables are influencing the result, satisfying transparency. However, SLIM as an ML technique is different from more notable ML approaches in its calculable properties (e.g. scalability). The methodology compared nine ML classifiers, including SLIM, on features that they claimed were typically accessed by police and judges (e.g. prior arrests and imprisonment history) to predict six offence categories: arrest; drug possession; and four types of violence. A limitation of this data collection method is that bias may be present in the data, as a proportion of it comes from police records, such as prior arrests. Despite this, the researchers reported a preference for SLIM over the other approaches, as it employs a simplified numerical scoring method, which is supported by its accuracy and interpretability.

Further promising results on machine versus human prediction have been reported in the literature. A study conducted by Singh, Jain, and Kumar (2017) examined the effectiveness of ML on parole release decisions for inmates seeking release from the State of New York. The researchers applied a neural network

algorithm to analyse 18,688 cases, with 70% used as training data, 20% as validation data and 10% as testing data. In all, 10 predictor variables were used, including age, sex and race. The researchers claimed that the accuracy of their model was calculated at 76.8%. The result, while auspicious, had several limitations; for example, the “black-box” program used hidden layers, and the algorithm was undisclosed (Waltl and Vogl 2018). An additional limitation was found from at least one attribute of “race”, which was certain to create inherent bias. Berk, Sorenson, and Barnes (2016) compared an ML program to a human-based assessment of released offenders and found that 10% of the offenders they deemed suitable for release had reoffended within two years, which was more accurate than the conventional assessment, which had a 20% reoffence rate. Despite the favourable results, a notable limitation in the methodology was observed, specifically regarding the high number of variables. The greater the number of variables or features applied, the greater the complexity and difficulty in interpretation (McCue 2014).

A more conventional decision-making approach to prediction was found in the literature. Lum, Boudin, and Price (2020) conducted a pilot study to determine the effectiveness of a predictive instrument and associated issues of bias and fairness in police charge categories. The predictive instrument used was the Arnold Public Safety Assessment (APSA), which can be described as a traditional statistical measure employing a six-point scale that incorporates various features (e.g. prior failure to appear in Court and prior convictions), resulting in a binary outcome. Its purpose is to determine if an accused person should be released or held after being charged with one or more offences.¹ Moreover, the aim of the instrument is to determine the probability of the accused, if bailed, reoffending and/or failing to appear in Court. The results from the scale assessment are then weighed against a decision matrix that was created for the pilot study. The researchers relied on data from criminal cases over a 12-month period (2016–2017) in a US criminal jurisdiction. Given that there were several items analysed by the authors, for brevity, there were two notable outcomes: 27% of bail cases were incorrectly assessed by the APSA, which would have resulted in unnecessary intervention and restrictions on those offenders, and 10 to 20% of offender cases analysed were found to have had the charges unsubstantiated by the Court (which corresponds with the data in NSW referred to earlier). By their admission, the authors acknowledged a small sample size and had not analysed potential biases in the data that might make an impact on judicial decisions and recidivism rates. Despite these limitations, the research by Lum et al. (2020) underscores the importance of ethics and morals in criminal justice, particularly in contexts where human assessments are fallible.

Schematizing Bail Decisions

It is intended that the decision-making process for bail in NSW criminal courts, demonstrated schematically and materially, can be replicated into a computerized model represented as a numerical score or measurable result. This proposition is supported in the literature, where an algorithm can be formulated to produce a decision, represented as a numerical score or probability. Contemporary

¹This is similar to police bail decisions in NSW.

algorithm-based risk assessments would need to apply specific measures or predictors, such as “criminal history” and “rearrest” (Berk and Elzarka 2020). However, caution was expressed when using predictor variables such as “past arrests” rather than “past convictions” as it can significantly affect prediction outcomes (Wykstra 2020).

A discussion on variables was considered in the literature. Stobbs et al. (2017) claimed that any algorithm coding for decision-making that has multiple variables can be produced. Stobbs et al. (2017) stated that 30 factors currently guide the NSW courts in sentencing legislation, although they reported that there are over 200 factors to consider; however, two key factors, criminal record and offences while on bail, were identified. A real-world model complements this discussion: Jung et al. (2020) designed and tested a model to determine whether accused persons should be granted bail, assessing two key features: age and prior failures to appear (Jung et al. 2020). After analysing over 100,000 pretrial detention cases, the authors claimed that their model would have allowed judges to detain one-third fewer people and would not have increased the risk of any individuals failing to appear at the next court hearing if they were bailed (Jung et al. 2020). A limitation in the research was found where a miscalculation may have occurred as to the reasons a bail authority released an individual due to inaccurate or missing information.

A separate research study examined the variables most prominent in failure-to-appear matters. Zettler and Morris (2015) applied a logistic regression model and analysed bail cases over a year to determine, from 19 predictors, such as criminal history and prior failure to appear, which factors were more likely to result in defendants failing to appear in Court. The researchers’ design variation included six models, five of which were grouped by race and gender to minimize bias, and a general model. The results demonstrated that the predictors of being male, impoverished and not having a prior felony charge² increased the chances of failure to appear. Despite the encouraging results, several limitations were noted in this research. First, the outcomes are only specific to that jurisdiction. Second, the monitoring of persons on bail by a government service is not applicable everywhere; government intervention in this instance could positively affect compliance. Third, the information on the subjects may have been limited, particularly the data on homeless status.

Despite the persuasive arguments regarding the number of predictor variables, the benefits of applying criminogenic factors in predictive models are evident. A meta-analysis conducted by Gendreau, Little, and Goggin (1996) on predictors of adult offending and recidivism concluded that some of the most significant criminogenic factors were age, gender, criminal history, pro-criminal associates, family circumstances and substance-related issues. Additionally, a synthesis of actuarial risk measures for recidivism revealed a higher correlation than personality measures, likely due to the heterogeneity of reoffending factors (Gendreau et al. 1996). In another meta-analysis, Grove et al. (2000) reviewed 136 studies on the topic of “clinical versus mechanical” or “human versus algorithm” prediction in the field of human health and behaviour. The results demonstrated that the algorithm’s prediction was more accurate than that of humans in predicting criminal behaviour. The meta-analysis reviewed studies dating from 1936 (parole success or failure) to

²A felony is similarly equated to an indictable offence in Australia.

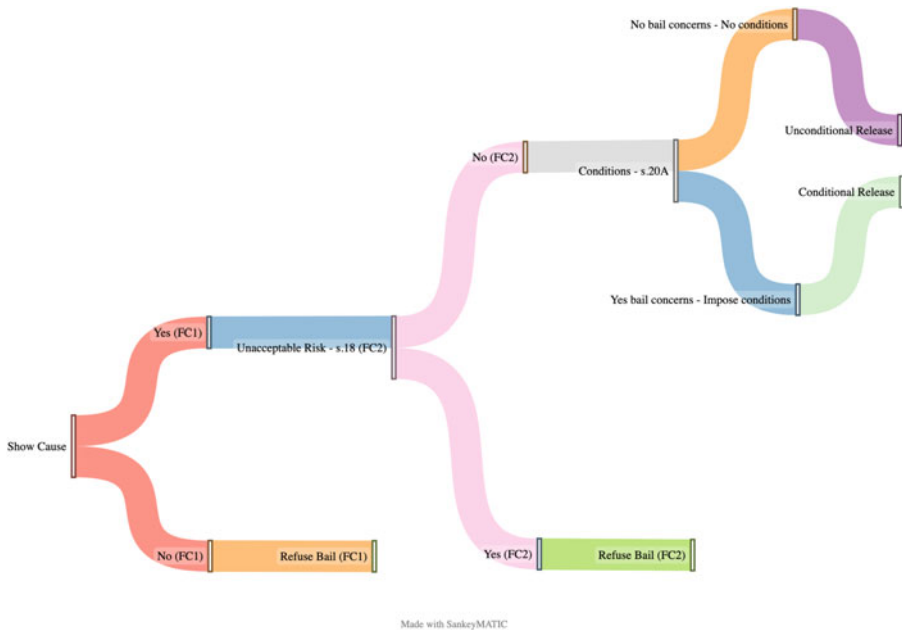


Figure 1. Schematic diagram *BAILgram* (created through SankeyMATIC; Bogart 2023). The process moves from left to right; the different colours differentiate each stage in the bail assessment; the channel over the top half is indicative of bail granted, while the bottom half is indicative of bail refused. The letter representations of “FC1” symbolizes Flow Chart 1: show cause requirement; and “FC2” symbolizes Flow Chart 2: unacceptable risk test. The colour intervals and literal notations signify a point where a decision is to be made in the same way as the bail legislation schema.

1988 (criminal behaviour), and it is noteworthy that forecasting criminality over that 52-year period did not utilize sophisticated computer software. Any comparison drawn from these outcomes to current forecasting should be made cautiously. Notwithstanding, statistical prediction was equal to or more favourable than human-made prediction.

Depicted in the Bail Act 2013 (NSW) by schematic flow charts is the process that ultimately leads to a binary or dichotomous decision of either “yes” (bail granted) or “no” (bail refused). The NSW Supreme Court succinctly expressed this:

The schema of the Act suggests a two-stage task in which the Court would first call upon the accused person to show cause why his or her detention is “not justified.” Subsection 2 of [section] 16A provides that if the accused person does show cause why his or her detention is not justified, the bail authority must make a bail decision by Division 2 of Part 3, which is the unacceptable risk test. That test applies to all offenses (*M v. R* 2015, 138).

The flowcharts referenced here have been reproduced as a Sankey diagram (Bogart 2023), which is aptly titled for this study as the *BAILgram* (see Figure 1).

B-LogR Algorithm Model as a Predictive Instrument to Decide Bail

Regression is essentially a statistical measure used to determine the effect of one or more independent variables on a dependent variable – it is a robust approach to predictive analysis (Pinder 2020). In criminological and justice domains, a B-LogR model is favoured, as the outcomes being sought are fundamentally dichotomous (Britt and Weisburg 2010); for example, bail granted versus bail refused, or an order to detain versus an order to release. B-LogR models show good predictive utility (Ngo, Govindu, and Agarwal 2015) and satisfy assessing or measuring its statistical significance on a dependent variable (Britt and Weisburg 2010). It is an effective statistical model when applied to observational data (Shmueli 2010) and in analysing binary outcomes (Zettler and Morris 2015). The multinomial model, as an extension, is effective because it can assess three or more categories (Britt and Weisburg 2010) and has also been said to exhibit low bias and “mean absolute error” (Wilkinson, Mamas, and Kontopantelis 2022). Nonetheless, these models have limitations. Presumably, bias can occur due to methodological and data errors (Kleinbaum et al. 2014), and small sample sizes can have a negative impact on results (Wilkinson et al. 2022). It is arguable, then, to suggest these limitations are hardly as acute as those observed in so-called “black-box” models, which, among other reasons, are maligned for lacking transparency and containing bias.

T-sC Algorithm Model as a Predictive Instrument to Decide Bail

The next approach is the T-sC. It is a supervised classification and predictive model (Wijenayake, Graham, and Christen 2018), which categorizes datasets into smaller subsets (Nasridinov, Ihm, and Park 2013) and ultimately provides a visual representation of the data, structured in a tree-like format (Wijenayake et al. 2018). This approach has been said to be favoured by social scientists in criminological domains for its ability to construct the probability of recidivism based on relevant factors (Lussier et al. 2019). Additionally, it resolves complexities in decision-making while simplifying interpretation and use (Lee, Liu, and Jin 2014), a characteristic also described by Kotsiantis (2013) as being intelligible. The T-sC also shows good predictive utility and accuracy (Ngo et al. 2015; Zeng et al. 2017) and was also touted as a preferable approach, as it corresponds well with flowcharts (Lee et al. 2014) and thus can be visualized. The T-sC is said to be a robust performer in predictive analysis, as it benefits from interpretability, is understandable and is less complicated (Kotsiantis 2013; Lee et al. 2014; Rutkowski, Jaworski, and Duda 2020). Accordingly, it is one model that has been selected and will be more suitable than other ML approaches in its relationship to the bail schema. Despite the many benefits of a T-sC, there are limitations. The literature suggests that models are not always interpretable, such as black-box models, and trade-offs are necessary for achieving both interpretability and accuracy (Zeng et al. 2017). Furthermore, increased model complexity can lead to negative outcomes (Kotsiantis 2013).

Summary

Uniformly, the literature on ML being applied in the criminal justice domain endorses logistic regression and tree classifiers for predictive modelling and

decision-making instruments. This, however, is not to discount other models that are effective in the criminological and justice domains; rather, the justification given for this was such that logistic regression is ideal for binary classification (Berk and Bleich 2013; Jung et al. 2020; Zeng et al. 2017) and mollifies ethical concerns beyond that of other techniques, such as neural networks (Attewell and Monaghan 2019) and random forests (Berk and Bleich 2013). Similarly, more modern developments in ML, such as neural networks, meant that its application as a tool was met with greater complexities than those of relatively less complex ones, like tree classifiers (Lotfi and Bouhadi 2022). The literature research also revealed various ML methods, where the predictive rigor of those selected was comparable – two of these being logistic regression and tree classifiers (Lussier et al. 2019; Ngo et al. 2015). Given these scholarly examples, which have sustained suitable ML methods for predictive modelling, specifically in the criminal justice domain, B-LogR and the T-sC are appropriate in laying the groundwork for AI-generated decision-making in bail.

Statistical Methods

There is an argument bound in historical practice that research and statistical methodology should be tested against theories and hypotheses and that the causal elements and statistical results should be determined by deduction. This study does not intend to form hypotheses and develop causation based on analyses underpinned by theoretical premises and model explanations; rather, it is centred on predictive modelling. It is imperative to provide the rationale for this inclination; however, it is constructive to differentiate “explanatory” from “predictive” modelling. Predictive modelling is the application of algorithmic modelled numerical data to forecast a future event or make a predictive observation (Shmueli 2010). Alternatively, explanatory modelling is a results-oriented undertaking that examines causal outcomes based on theories and hypotheses (Shmueli 2010). While explanatory modelling contains measurable data, it becomes misrepresented by the theory underpinning it, leading to inefficacy in explaining explainable phenomena (Shmueli 2010). Notwithstanding, a key requirement for predictive accuracy in the criminal justice domain is that the future has representations of the past, and future predictions are dependent on mathematical rigor obtained through superior data collection and analysis (Berk and Bleich 2013).

A contemporary approach to measuring the risk of reoffending involves ascertaining the relationship between predictor variables and their impact on a target variable (Berk and Elzarka 2020). Similarly, this study will investigate whether any relationship exists between the categorical variables to predict the probability that an accused, if released on bail, will reoffend. Purposed as reliable measures of the performance of an ML model will be the receiver operating characteristic (ROC) curve and classification table (Attewell and Monaghan 2019), for their predictive vigor was reasonably justified by the account of analogous classification tables that conceivably lead toward a lower chance for unfairness (Berk and Elzarka 2020).

After, or working diachronically with, the predictive model approach is exploratory data analysis (EDA). EDA is not wholly defined within the literature but is characteristic of analysis methods that precede traditional methods – upon which analysis can be drawn – where hypotheses are not a foundational requirement in a

Table 1. Number of defendants in the respective New South Wales courts corresponding to data collation ($n = 101$)

Court type	Number of cases/defendants
Court of Criminal Appeal	34
Supreme	60
District	7

study (Shmueli 2010). The rationale for non-traditional methodologies is sustained conditionally, as once the outcomes were in an analysable format, it became evident that there would be a disconnect with previously formulated hypotheses on causation for the bail decisions. Supported in the literature, sizeable datasets are presented with correlational and pattern complexities, which become difficult to hypothesize (Shmueli 2010).

Further sustaining the shift from traditional statistical methods is the concern about the probability value. This discussion on using the p value, as it is commonly referred to, is extended to significance testing. The p value is usually accompanied by arbitrary indicators of 0.05 or 0.01 (Dahiru 2008). The preferred measurement criterion is to apply and rely on “confidence intervals” for their superiority over hypothesis testing, for example, in study replication (Dahiru 2008).

This paper’s groundwork is based on the development of a working prototype to predict bail decisions using actual data – specifically, to forecast whether a defendant should be granted or refused bail based on nine distinct predictor variables. Considering the contentious yet compelling scholarly rationale, this study will depart from tradition and refrain from formulating hypotheses, although a reasonable statistical approach will be presented about the predictive model.

Data Evaluation and Metrics

A collation of 101 bail case narratives across various levels of the NSW criminal courts formed the dataset, accessed from the NSW Caselaw website (<https://www.caselaw.nsw.gov.au/>). Caselaw is an open-source, publicly accessible website with written decisions on nominated cases in most jurisdictions of the NSW law courts. The search parameters contained “bail”, the chosen four court levels of “local”, “district”, “supreme” and “Court of Criminal Appeal (CCA)”, and the period parameter from “2015” to “2023”, as 2015 was the year of the last legislative amendment. The search parameters made available more than 600 cases; however, this was eventually filtered down to 101 case narratives based on detail and relevance from the District and Supreme Courts, as well as the CCA (see Table 1). It is noted that some of the cases contained two or more multiple parties heard at the same time (co-defendants), although each defendant was counted as one case and assessed individually, as the decisions and factors contributing to those cases were dealt with separately by the respective courts.

Data were manually recorded corresponding to each variable. For instance, where a written judgment stated that a defendant had one prior failure to attend

Table 2. Classification table exemplar for bail decisions

	Bail granted (actual decision)	Bail refused (actual decision)
Bail granted (predicted decision)	True positive (TP)	False positive (FP)
Bail refused (predicted decision)	False negative (FN)	True negative (TN)

court, this was subsequently recorded under the predictor “failure to appear/flight risk” as “1”. If a judgment did not contain the information to satisfy each of the predictors, it was disregarded. Each predictor variable is recorded commensurate to the information in the Model Predictor Information Table (see Appendix). All variables were binary coded, with values of either “0 = no” or “1 = yes”, except for “criminal history” and “seriousness of offence(s)”, which required more specific coding. Under the class “seriousness of offence(s)”, if viewed on a scale, it can vary depending on factors associated with harm to victims. Therefore, seriousness was rated as the maximum penalty that could be imposed under the sentencing law for the index offence. Guided by domain experience in sentencing procedures and administration, seriousness was gauged by the penalty and the hierarchy of the court. For example, the three-year penalty in sentencing is a significant marker, as a penalty of this degree can only be determined by superior courts. A three-year or more custodial sentence requires the NSW State Parole Authority to determine the schedule and conditions of an offender’s release. “Show cause” under the legislation is a “yes” or “no” outcome and is therefore coded as “1” and “0”, respectively. “Show cause” data input determines if the defendant has shown cause why their detention is not justified, which resembles the NSW bail legislation but was varied, giving greater weight to certain offences in the categories of sex crimes and domestic/family violence. It is important to note that the interpretation of this class is transposed for the two ML techniques.

One of the determining factors for bail in the NSW legislation refers to “if” a defendant was convicted, whereby the emphasis on *if* is a subjective assessment made by a magistrate or judge on the probability of a conviction. To minimize this subjectivity (where a human is tasked with considering numerous factors and predicting the probability that a defendant will or will not comply with conditional release), the 24 factors were condensed into nine predictor variables that comprise the prototype, titled *Bail-14*, making the predictors less ambiguous. Attributes such as gender, age and race have intentionally been omitted, which, from the literature, can create biases in the results, and from a logical and moral standpoint, the weight of those attributes should not give cause to determine a defendant’s liberty.

Classification Metrics

Table 2 is an exemplar of a classification table with specific reference to the bail decision categories. The letters are literal notations of true positive, true negative, false negative and false positive (Larner 2021). The vertical columns represent the actual or true classifier, and these are put against the outcomes from a predicted classifier in the horizontal columns (Larner 2021).

Table 3. Error-based measures and information-based measures linked to bail decision

Measures	Equation
TPR: the probability of a correct classification of Granted	$TP/(TP + FN)$
TNR: the probability of a correct classification of Refused	$TN/(FP + TN)$
FPR: the probability of incorrect classifications – Granted being Refused	$FP/(FP + TN)$
FNR: the probability of incorrect classifications – Refused being Granted	$FN/(FN + TP)$
PPV: the probability of compliance given an outcome of Granted	$TP/(TP + FP)$
NPV: the probability of non-compliance given an outcome of Granted	$TN/(TN + FN)$

Note: TPR, true positive ratio; TP, true positive; FN, false negative; TNR, true negative ratio; TN, true negative; FP, false positive; FPR, false positive ratio; FNR, false negative ratio; PPV, positive predictive value; NPV, negative predictive value.

Error-based measures are summarized in the following and listed in Table 3. Each measure is categorized according to its application in bail decisions. Equations for each measure are also stated in Table 3. The true positive ratio (TPR) and true negative ratio (TNR), respectively, are indicative of either a positive classification or a negative classification in a specified category. The false positive ratio (FPR) and false negative ratio (FNR) are probabilities of incorrect classifications, generally speaking, a finding of incorrect when it is correct, or conversely, a correct finding when it is incorrect. Information-based measures include the positive predictive value (PPV) and the negative predictive value (NPV), which are, respectively, the probabilities of a finding in a specified category being correctly or incorrectly predicted. Outcomes for all ratios and values are determined numerically between 0 and 1 (Larner 2021).

Nine variables in the *Bail-14* prototype are derivatives of the 24 factors listed under Section 18 of the Bail Act 2013 (NSW). It was deemed that some of those listed factors overlapped or were similar in context, and for the formulation of *Bail-14*, those factors were either merged or omitted. Also, some items appear subjective, and their evaluative cogency is consequently limited. Given these factors, the nine predictors applied to the *Bail-14* model are listed in Table 4.

Demographic identifiers, while relevant in certain assessments, have been shown to create issues with fairness (among other ethical principles) in the justice and law domains. For these reasons, demographic identifiers will not be used. In saying this, the Bail Act 2013 (NSW) does not contain demographic criteria.

B-LogR

This is a parametric statistical approach that ultimately produces a dichotomous outcome. It is a cogent predictive measure in binary classification for values of 0 or 1, true or false, or yes and no (Zaidi and Al Luhayb 2023), suitable for predictive modelling where variables of consequence need to be ascertained (Barabas et al. 2018). When considering this approach from an ethical standpoint in comparison to other ML approaches, the B-LogR model is understandable (Zaidi and Al Luhayb 2023), which encompasses ethical principles such as interpretability,

Table 4. Nine predictor variables of the *Bail-14* model

1. Show cause
2. Criminal history
3. Seriousness of the offences(s)
4. History of violence
5. Bail non-compliance (whether under the Bail Act 2013 (NSW) or other jurisdiction)
6. History of non-compliance (with court-issued orders)
7. Pro-criminal associations
8. Threat/danger to the victim(s), public or others
9. Failure to appear/flight-risk

Note: See the Appendix for more specific information on the nine predictors.

explainability and transparency. Notwithstanding the ethical dilemmas that may arise from prediction as an undertaking in itself, an ML-driven regression measure is applied to determine and explain the relationship between the predictors and the dependent variable and to evaluate the predictive power of this relationship (Alvo 2022).

The data were split into several ratios to account for overfitting and to determine model accuracy (Attewell and Monaghan 2019; Berk and Bleich 2013; Martino 2019). A multivariate multiple regression statistical method was applied. The reason for this method is due to the two dependent variables: the response variable (y) has two qualitative values (Johnson and Wichern 2013) – in this instance, granted and refused – and, as such, the actual court decision from each bail case is the target variable.

The model applied here is an archetype of a logistic regression algorithm (Jankovic 2021; Ngo et al. 2015; Pennsylvania State University 2018) and is inspired by a model from a scholarly source (Jung et al. 2020):

$$p(y) = \frac{1}{1 + \exp(a + b^1c^1 + b^2c^2 + b^3c^3 + \dots b^9c^9)},$$

where p is the probability of success (bail granted = y), a is the intercept, b are the coefficients corresponding to a recorded instance of the predictor c and refers to the binary instance of 0, “no”; 1, “yes”.

As listed in Table 3, each predictor variable is accompanied by its coefficient, as demonstrated in the equation above: show cause (c^1); criminal history (c^2); seriousness of offence(s) (c^3); history of violence (c^4); bail non-compliance (whether under the Bail Act 2013 (NSW) or other jurisdiction) (c^5); history of non-compliance (with court-issued orders) (c^6); pro-criminal associations (c^7); threat/danger to the victim(s), public, or others (c^8); and failure to appear/flight risk (c^9).

A code-based program assigned the designation of the codes and values, as described by Colectica (2024), which is a Microsoft Excel add-in that supports

coding variables. Another Excel add-in, Real Statistics Using Excel Resource Pack (Zaiontz 2024), was applied, which enables logistic regression functions while facilitating the coding of categorical variables. These packages analysed the data from the 101 bail case narratives to determine the probability of success or failure for a defendant being granted or refused bail based on the nine predictors listed in Table 4.

Acceptably applying domain knowledge in risk assessment and jurisprudence (Alikhademi et al. 2022), three or more convictions were a satisfactory threshold of an individual's historical offending. The "seriousness of offence(s)" classification was based on the current penalty range for the index offence by the NSW sentencing legislation: equal to or less than three years' imprisonment (1, low); greater than three years/not more than 10 years' imprisonment (2, moderate); and greater than 10 years' imprisonment (3, high). More specific details can be found in the Appendix.

While the threshold is gauged or presumed by the intentions or purpose of the model, it was determined that 0.4 was a suitable marker, given that the accuracy was at its highest. Consequently, the classification "cut-off" or "threshold" was set at 0.4. When determining whether to grant or refuse bail, the chance bet is not a suitable option; rather, a definitive metric is needed. A defendant who scores equal to or below 0.4 would be refused bail, while one who scores equal to or greater than 0.41 is indicative of a defendant being granted bail. Moreover, the literature suggests that a threshold of 0.5 is an arbitrary marker, as it is seemingly not any more definitive than chance (Attewell and Monaghan 2019; Chan 2004; Clipper 2016; Helmus and Babchishin 2017). Accuracy is calculated by the addition of true positives and true negatives divided by the total number of bail cases tested (Larner 2021). The ROC curve charts the effectiveness of the model in predicting a target, discernible by the curved line that follows the y -axis to the top left corner and then turns to move parallel to the x -axis (Attewell and Monaghan 2019; Clipper 2016).

Several models were tested, aptly named after the split ratios: *Model 61-40* and *Model 51-50*. Additionally, the models were also tested on their performance by removing two predictor variables, selected on domain knowledge in addition to the literature on recidivism (Jung et al. 2020; Skeem and Lowenkamp 2020), and these sub-models are labelled as *-CRIM50* and *-SoO50*.

T-sC

Otherwise referred to as a decision tree, the T-sC is a non-parametric approach derived from the *Bail-14* dataset and is herein referred to as "Bail-Tree". Produced using the open-source software RapidMiner-Studio (RapidMiner 2024), the data applied were imported from Excel, much like B-LogR, although with two distinct exceptions. The first exception is that the target variable does not require coding, as the polynomials of "granted" and "refused" serve as deliberate set parameters. The second exception is that the "show cause" outcome is transposed, for in this process, the decision is determined by the defendant having to show cause why detention is not justified, which is the same as asking what legal argument the defendant can submit to justify their liberty, for example, not presenting any risk to the victim or

Table 5. *Bail-14* pseudocode exemplar to demonstrate the six simplified commands or syntax to build the tree-structured classifier at a depth of eight

1. Select bail cases.
2. For each predictor P , find the normalized information gain ratio from splitting P
3. Let P best be the predictor with the highest normalized information gain
4. Create a decision node that splits on P best
5. Recur on the sublists obtained by splitting on P best, and add those nodes as children of the node $(P_i, (P_{ii}, P_{iii} \dots P_{viii}))$
6. Those will be children of node P which has the highest information gain ratio.

community (whereas the B-LogR interrogates the specific offence for which the defendant is being charged similar to the legislative framework, e.g. was the alleged offence committed while on bail or parole). As noted earlier, adjustments to this classification were made to give greater weight to categories under sex crimes and domestic/family violence. The target variable was labelled “actual decision”.

RapidMiner-Studio does not exclusively identify what algorithm model is specifically applied under the “Decision Tree” operator, although it closely replicates the perennial algorithms ID3 and C4.5 (see Quinlan 1986; Ramakrishnan 2009). Bearing a closer resemblance to the C4.5 algorithm, Bail-Tree initially applies a top-down approach (Ramakrishnan 2009), meaning that the first or root node is determined to be the most relevant predictor and subsequent nodes are determined accordingly until the last two leaves display the outcomes. More specifically, it systematizes the strength or value of all predictors selectively through data computation until it reaches a juncture, where the return path is again calculated by a bottom-up process that ultimately selects the most relevant predictor at the root to then form the branches and leaves by “splitting” according to their relevance or strength (Quinlan 1986; Ramakrishnan 2009). In its more practical utility, the schematic flow in the Bail Act 2013 (NSW) is also a top-down assessment upon which a bail authority is to reference in that decision-making process. A utility of the T-sC was its ability to simulate analytical decisions (Wijenayake et al. 2018).

A split was applied to the training (0.6/60%) and test data (0.4/40%), and to build the model, a “stratified sampling” option was applied as it was considered the most suitable for binomial calculation; “gain ratio” was the criterion deemed most suitable for splitting (RapidMiner 2024). The confidence interval was set at 0.1, the minimal gain at 0.01, and the minimal leaf size was set at 2 (size for split set at 4). The integers ranged between 0 and 3, and polynomials were granted/refused and yes/no. Table 5 implies the “pseudocode” exemplar for this process.

Results

B-LogR

Model 61-40

Training was conducted on 61 randomly selected cases and, subsequently, the remaining 40 cases were used as test data. The classification performance in Table 6

Table 6. Classification table for *Model 61-40*

Predicted class	Actual class	
	Granted	Refused
Granted	19	7
Refused	2	12

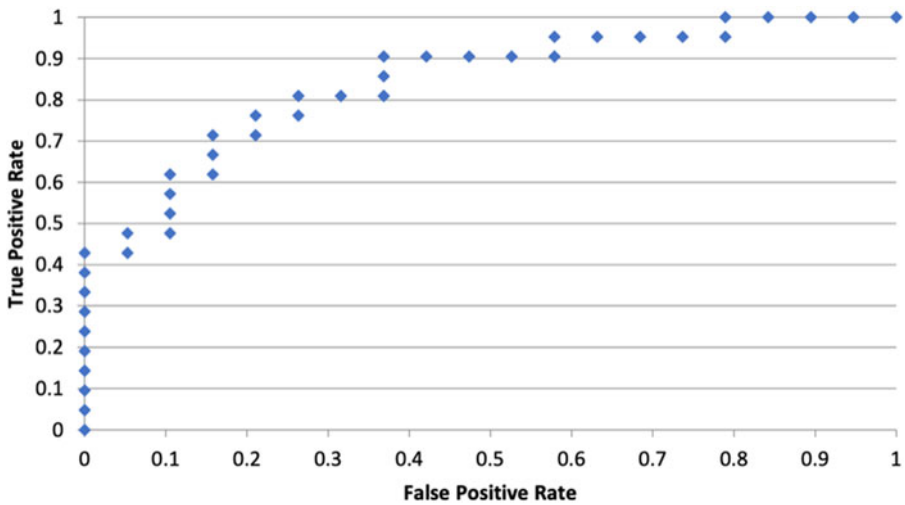


Figure 2. Receiver operating characteristic curve for *Model 61-40* (area under the curve 0.845, 95% confidence interval).

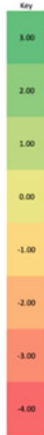
indicates an overall accuracy of 0.775 (78%); there were 19 (0.90) successful observations and 12 (0.63) failed observations. Figure 2 displays the applicable ROC curve (area under the curve (AUC) 0.845) and is indicative of moderate to strong model accuracy. Statistically assessed using the conventional predictive value ($p < 0.05$), the model was rendered significant ($p = 0.02$).

Turning to information-based and error-based measures, the TPR suggests the probability of correct classification of Granted is 90% (0.90), while the TNR suggests the probability of correct classification of Refused is 63% (0.63). The FPR at 36% (0.36) and FNR at 9% (0.09) refer to the probability of incorrect classifications. The PPV (0.73) and NPV (0.86) are indicative that the probability of compliance given an outcome of Granted is 73%, and the probability of non-compliance given an outcome of Granted is 86%.

Table 7 shows in a variance-covariance matrix that “criminal history” and “failure to appear/flight risk” resulted in a negative correlation ($-1.8E+08$). “Bail non-compliance” featured prominently: a negative correlation with “history of violence” (-0.09) and similarly with “pro-criminal associations”, with a positive

Table 7. Variance–covariance matrix for *Model 61-40*

	Criminal history	Seriousness of Offence(s)	History of violence	Bail non-compliance	Non-compliance with other orders	Pro-criminal associations	Threat towards victim/others	Failure to appear/ flight-risk	Show Cause Offence	Actual Decision
Criminal history	1.80E+08	-1.52	-3.26	0.34	2.26	-0.98	-1.72	-1.80E+08	1.19	0.96
Seriousness of Offence(s)	-1.52	0.33	0.31	-0.16	-0.24	-0.06	0.18	0.54	-0.24	-0.05
History of violence	-3.26	0.31	1.25	-0.09	-0.76	0.85	0.27	0.28	-0.34	-0.89
Bail non-compliance	0.34	-0.16	-0.09	0.56	-0.22	0.09	-0.14	-0.11	-0.08	0.18
Non-compliance with other orders	2.26	-0.24	-0.76	-0.22	2.64	-1.85	-0.06	-0.51	-0.24	0.68
Pro-criminal associations	-0.98	-0.06	0.85	0.09	-1.85	2.70	-0.04	-0.63	0.03	-1.25
Threat towards victim/others	-1.72	0.18	0.27	-0.14	-0.06	-0.04	1.50	0.28	-0.66	-0.21
Failure to appear/ flight-risk	-1.80E+08	0.54	0.28	-0.11	-0.51	-0.63	0.28	1.80E+08	-0.34	0.48
Show Cause	1.19	-0.24	-0.34	-0.08	-0.24	0.03	-0.66	-0.34	1.52	0.06
Actual Decision	0.96	-0.05	-0.89	0.18	0.68	-1.25	-0.21	0.48	0.06	1.69

**Table 8.** Classification table for *Model 51-50*

Predicted class	Actual class	
	Granted	Refused
Granted	20	8
Refused	4	18

correlation (0.09). “Show cause” was negatively correlated with “bail non-compliance” (−0.08) and positively correlated with “pro-criminal associations” (0.02).

Model 51-50

The training was conducted on 51 randomly selected cases. Subsequently, test data were taken from the remaining 50 cases, as displayed in Table 8. Overall accuracy was 76% (0.76), with 20 (0.83) successful observations and 18 (0.69) failed observations. Figure 3 displays the applicable ROC curve (AUC 0.845) determining accuracy. At 0.845, the model represents moderate to strong accuracy. The predictive value was calculated at a significance level of 0.01, which determined model significance ($p < 0.05$).

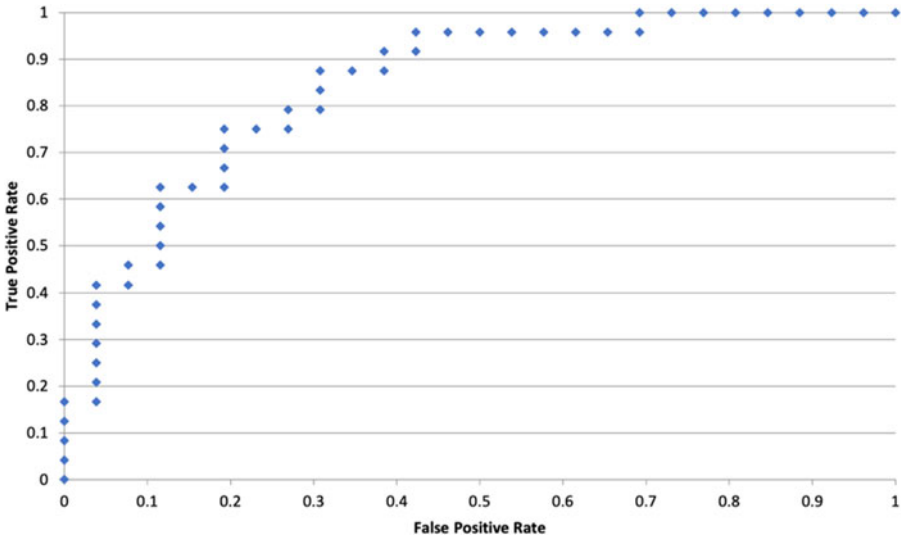


Figure 3. Receiver operating characteristic curve for *Model 51-50* (area under the curve 0.845, 95% confidence interval).

As for information-based and error-based measures, the TPR suggests the probability of correct classification of Granted is 83% (0.83), and the TNR suggests the probability of correct classification is 69% (0.69). The FPR at 31% (0.31) and the FNR at 16% (0.16) refer to the probability of incorrect classifications. The PPV value of 0.71 and the NPV value of 0.82 indicate the probabilities of “compliance” given the outcome of Granted (71%) and “non-compliance” given the outcome of Granted (82%), respectively.

Table 9 shows in a variance–covariance matrix that a positive correlation is observed between “seriousness of offence(s)” with “history of violence” (0.036) and “failure to appear/flight risk” (0.049), while “seriousness of offence(s)” has a negative relationship with “bail non-compliance” (–0.097), “non-compliance with other orders” (–0.040) and “pro-criminal associations” (–0.009). A positive correlation was observed between “history of violence” and “failure to appear/flight risk” (0.030).

Taking into consideration the literature on evaluating predictive modelling performance (Jung et al. 2020; Skeem and Lowenkamp 2020), it was warranted to measure what relevance the prominent two predictors had on the performance of *Model 51-50*, namely, “criminal history” and “seriousness of offence(s)”. To measure this, a datapoint comparison was undertaken using the ROC values of the TPR and FPR from *Model 51-50*, with the ROC values of the two sub-models, following the deletion of the relevant classification.

In Figure 4, the two sub-models are denoted as “-*CRIM50*” and “-*SoO50*”. Datapoints and the models’ performance are noted after the respective predictor class was removed, demonstrating little variation when interchanging these two predictor classes (see Tables 10 and 11).

Table 9. Variance-covariance matrix for Model 51-50

	Criminal history	Seriousness of Offence(s)	History of violence	Bail non-compliance	Non-compliance with other orders	Pro-criminal associations	Threat towards victim/others	Failure to appear/tight-risk	Show Cause	Actual Decision	Key
Criminal history	3.16	-0.25	-0.83	0.34	-0.21	-0.60	-0.30	-0.30	0.10	-0.24	3.5
Seriousness of Offence(s)	-0.25	0.12	0.04	-0.10	-0.04	-0.01	0.06	0.05	-0.06	0.02	3
History of violence	-0.83	0.04	0.40	-0.07	0.15	0.14	-0.05	0.03	-0.20	-0.14	2.5
Bail non-compliance	0.34	-0.10	-0.07	0.39	0.06	-0.28	-0.14	-0.13	0.09	-0.06	2
Non-compliance w/ other orders	-0.21	-0.04	0.15	0.06	1.25	-0.86	-0.25	-0.08	0.17	-0.29	1.5
Pro-criminal associations	-0.60	-0.01	0.14	-0.28	-0.86	1.62	0.51	-0.05	-0.32	0.16	1
Threat towards victim/others	-0.30	0.06	-0.05	-0.14	-0.25	0.51	1.25	-0.13	-0.50	-0.19	0.5
Failure to appear/tight-risk	-0.30	0.05	0.03	-0.13	-0.08	-0.05	-0.13	0.81	-0.15	-0.13	0
Show Cause	0.10	-0.06	-0.20	0.09	0.17	-0.32	-0.50	-0.15	1.03	0.25	-0.5
Actual Decision	-0.24	0.02	-0.14	-0.06	-0.29	0.16	-0.19	-0.13	0.25	0.94	-1

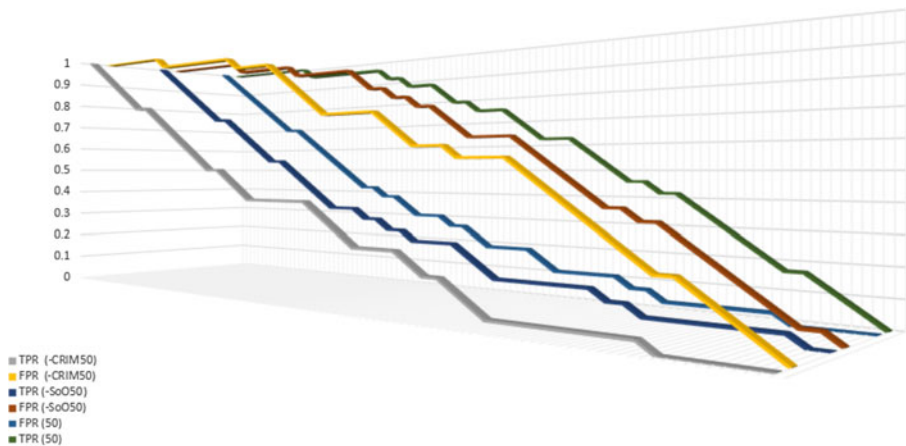


Figure 4. Model performance measured by the true positive ratio (TPR) and false positive ratio (FPR) of sub-models -CRIM50 and -SoO50 when the respective predictor class was removed.

As referred to previously, the full model determined the TPR at 83% (0.83) and the FPR at 31% (0.31). After “criminal history” was removed, the sub-model determined the TPR as 91% (0.91) and the FPR as 42% (0.42). Then, “seriousness of offence(s)” was removed, and the sub-model determined the TPR as 83% (0.83) and the FPR as 35% (0.35).

Table 10. Classification table for sub-model -*CRIM50*

Predicted class	Actual class	
	Granted	Refused
Granted	22	11
Refused	2	15

Table 11. Classification table for sub-model -*SoO50*

Predicted class	Actual class	
	Granted	Refused
Granted	20	9
Refused	4	17

Table 12. Classification table for *Model T-sC* (accuracy)

Predicted class	Actual class		Class precision (%)
	Granted	Refused	
Granted	12	5	70.59
Refused	6	17	73.91
Class recall (%)	66.67	77.27	

T-sC

Overall accuracy was 72.5%, and classification error 27.5%. As shown in Table 12, the TPR value is 0.67, suggesting the probability of a correct classification of Granted is 67%; the TNR value is 0.77, suggesting the probability of a correct classification of Refused is 77%.

For information-based and error-based measures, the PPV and NPV were 0.71 and 0.74, respectively, and are indicative that the probability of “compliance” given an outcome of Granted is 71% and the probability of “non-compliance” given an outcome of Granted is 74%. The FPR and FNR measure the probability of incorrect classifications: the probability of “granted being refused” was 22% (FPR value 0.22) and, conversely, “refused being granted” was 33% (FNR value 0.33).

Figure 5 displays an ROC curve, which is a measure of the T-sC model’s performance at a tree depth of “seven”. Denoted by the red line (AUC 0.702), the model is moderately accurate. When selected at tree depths of “eight” and “nine”, the overall accuracy was 72.5%. When there was no specified tree depth (indicated by -1), the accuracy was maintained at 72.5%. Figure 6 is the T-sC model descriptor for the unspecified depth, and Figure 7 is the T-sC visual output for the tree depth at “eight”.

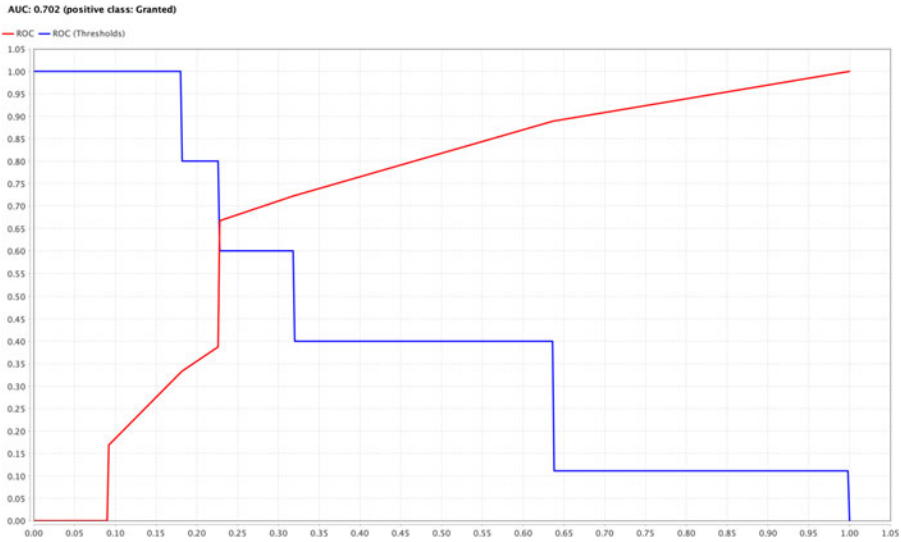


Figure 5. Receiver operating characteristic (ROC) curve for the tree-structured classifier model (area under the curve; AUC 0.702). The red line denotes the standard plots on the x-axis and y-axis and the blue line denotes the ROC threshold (values on y-axis are reversed). Graph output is a feature of the “Performance” classification parameters by RapidMiner (2024).

The trade-off in attempting to increase overall accuracy affected the true-positive and true-negative values. For example, when the minimal gain parameter was adjusted to 0.05, it resulted in an increased TPR (0.77) and TNR (0.54), although the overall accuracy was reduced to 65%.

Upon examination of the *Bail-14* predictive model’s error-based and information-based results, it is meaningful to undertake a comparison of the statistical data on bail-related matters over the period 2015 to 2023. The data comparisons relied on have been extracted from the NSW Bureau of Crime Statistics and Research (2023). Figure 8 graphs the decisions of the criminal courts for bail in NSW (excluding the Children’s Court) proportionate to defendants who were “bail refused”, and those defendants “bail granted” between 2015 and 2023. Notably, there has been a gradual increase in granted decisions since 2018, although the number of refused decisions has remained stable.

A comparison of “bail granted” numbers compared to the total number of matters with a bail status at finalization is presented in Figure 9. An inverse relationship is notable: when bail matters increase, the number granted decreases, and when bail matters decrease, the number proportionate to that for granted increases. Anecdotally, the inconsistency in the inverse relationship observed during the three years from 2019 to 2022 could be attributed to the COVID-19 pandemic, which necessitated modifications to court decisions due to unprecedented circumstances (NSW Bureau of Crime Statistics and Research 2023). However, this cannot explain the other inverse inconsistency between 2015 and 2019.


```

Seriousness of Offence(s) > 2.500
| Show Cause (Y/N) = No: Refused {Granted=0, Refused=5}
| Show Cause (Y/N) = Yes
| | Criminal history > 2.500: Refused {Granted=1, Refused=8}
| | Criminal history ≤ 2.500
| | | Criminal history > 1.500: Granted {Granted=4, Refused=2}
| | | Criminal history ≤ 1.500
| | | | Bail non-compliance > 0.500: Refused {Granted=0, Refused=3}
| | | | Bail non-compliance ≤ 0.500
| | | | | History of violence > 0.500: Refused {Granted=0,
Refused=2}
| | | | | History of violence ≤ 0.500
| | | | | Pro-criminal associations > 0.500: Granted
{Granted=3, Refused=2}
| | | | | Pro-criminal associations ≤ 0.500: Refused
{Granted=2, Refused=3}
Seriousness of Offence(s) ≤ 2.500
| Pro-criminal associations > 0.500
| | Seriousness of Offence(s) > 1.500
| | | History of violence > 1.500: Granted {Granted=4, Refused=2}
| | | History of violence ≤ 1.500
| | | | Failure to appear/Flight Risk > 0.500: Refused {Granted=0,
Refused=3}
| | | | Failure to appear/Flight Risk ≤ 0.500
| | | | | Show Cause (Y/N) = No: Granted {Granted=2, Refused=0}
| | | | | Show Cause (Y/N) = Yes: Refused {Granted=1, Refused=2}
| | | | | Seriousness of Offence(s) ≤ 1.500: Granted {Granted=2, Refused=0}
| | | | | Pro-criminal associations ≤ 0.500: Granted {Granted=8, Refused=2}

```

Figure 6. Tree-structured classifier model descriptor results based on *Bail-14* data. Y, yes; N, no.

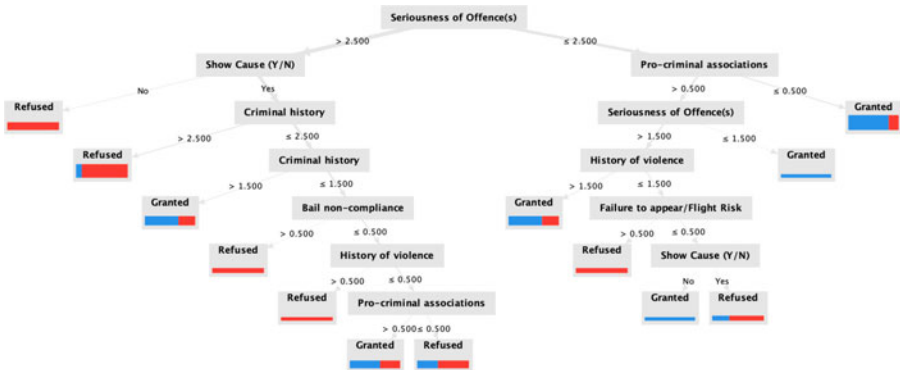


Figure 7. Screenshot of the tree-structured classifier output at a tree-depth of “eight” from *Bail-14* data. Statistical data comparison of NSW Bureau of Crime Statistics and Research (2015–2023) and *Bail-14* predictive model output. Y, yes; N, no. For predictor relevance order based on this figure, see Table 14.

Error-based measures and information-based measures from the two regression models 51-50 and 61-40, and tree-classifier model T-sC, are displayed in Figure 10. The box plots are mostly concentrated in the same areas for each model, although the T-sC does not demonstrate the same consistency as the two regression models.

To draw a reasonable comparison between the three models and recent bail statistics, particularly in the categories of “granted”, “refused” and “breach of bail”,

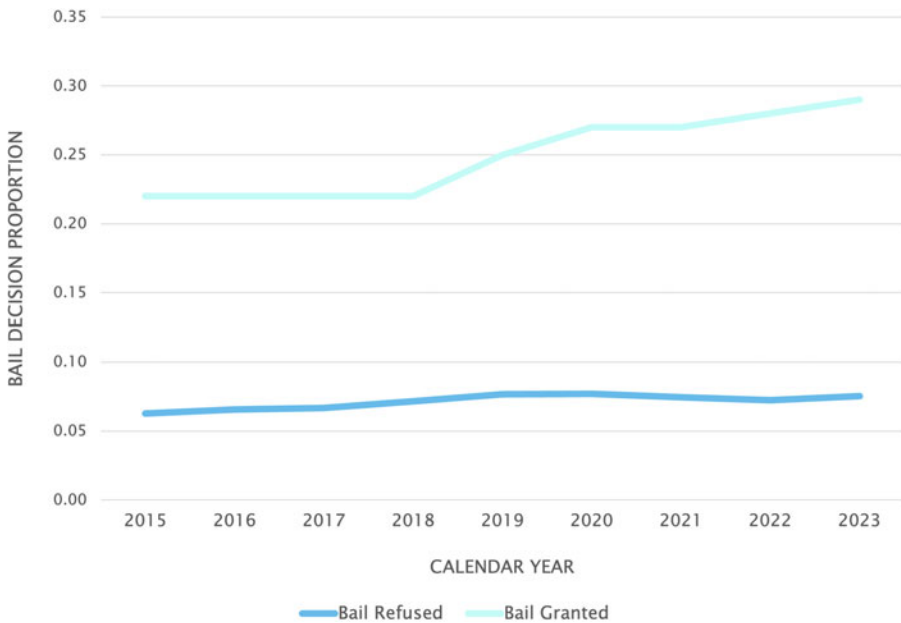


Figure 8. Bail decisions proportionate to the total number of bail matters at finalization. Raw numbers were extracted from NSW Bureau of Crime Statistics and Research (2015–2023) and calculated as a proportion to the total number of defendants who had bail matters before all adult courts in New South Wales over the period 2015–2023. Note that “finalization” refers to a defendant’s bail status at their final court appearance.

probability distributions were calculated and then compared to the error- and information-based measures. Recalling that an inverse relationship was previously observed between the two variables, it was therefore pertinent to base the predictive measure outcomes within the same period. The following equation is applied to determine the probability distributions:

$$f(x) = P(X = x),$$

where P is the probability, X is the random variable, and x is the mean (Pennsylvania State University 2024). Table 13 lists the results from this equation where “Success” equates to the probability of “bail granted” and “Failure” equates to the probability of “bail refused”.

Figure 11 details the values relative to the three calendar years from Table 13. It is apparent that the majority of the calculated error-based and information-based values within each year share a uniformity with Success (Granted) and Failure (Refused).

Lastly, it is of interest to revisit the earlier anecdote regarding the inverse relationship between “bail granted” and the “total numbers of bail matters finalized”, with the focus being on a three-year comparison of “bail granted”, “breach of bail”, and the PPV and NPV. As shown in Figure 12, the rates of defendants being granted and breaching their bail are relatively balanced. PPV and NPV data from *Bail-14*, when measured against those categories – the probability of

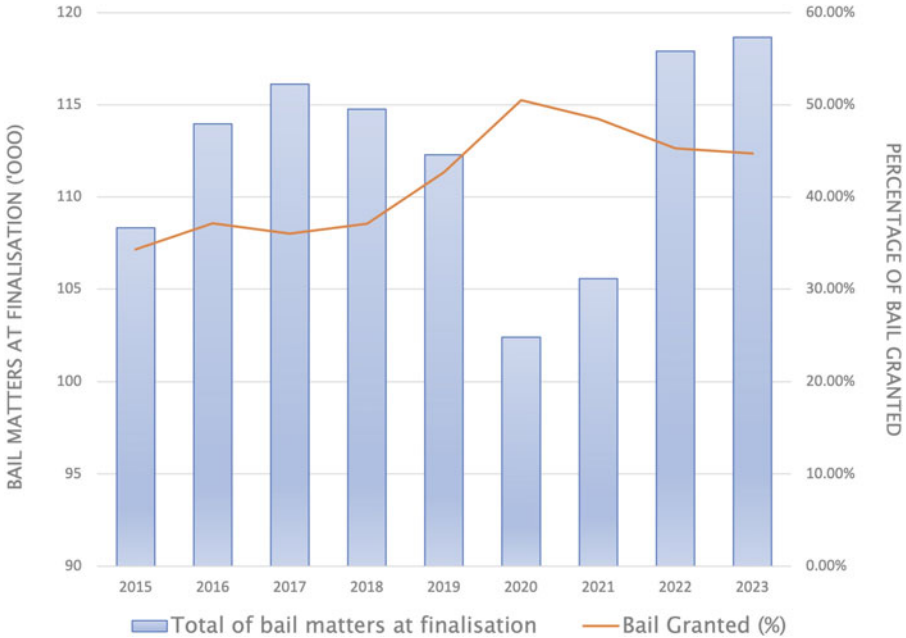


Figure 9. Bail status at finalization – all defendants compared to percentage of defendants granted bail. Data extracted from NSW Bureau of Crime Statistics and Research (2023).

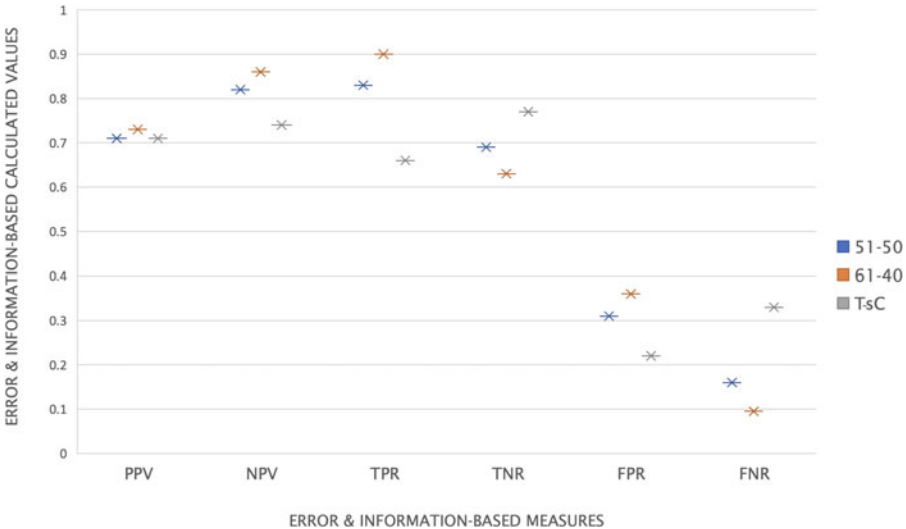


Figure 10. Error- and information-based measures of the two full regression models (51-50 and 61-40) and the tree-structured classifier (T-sC) model. PPV, positive predictive value; NPV, negative predictive value; TPR, true positive ratio; TNR, true negative ratio; FPR, false positive ratio; FNR, false negative ratio.

Table 13. Success–failure values comparison with error-based and information-based measures by year

	2015	2016	2017
Success	0.74	0.74	0.74
Failure	0.93	0.92	0.92
PPV	0.71	0.73	0.71
NPV	0.82	0.86	0.74
TPR	0.83	0.90	0.66
TNR	0.69	0.63	0.77
FPR	0.31	0.36	0.22
FNR	0.16	0.10	0.33

Note: PPV, positive predictive value; NPV, negative predictive value; TPR, true positive ratio; TNR, true negative ratio; FPR, false positive ratio; FNR, false negative ratio.

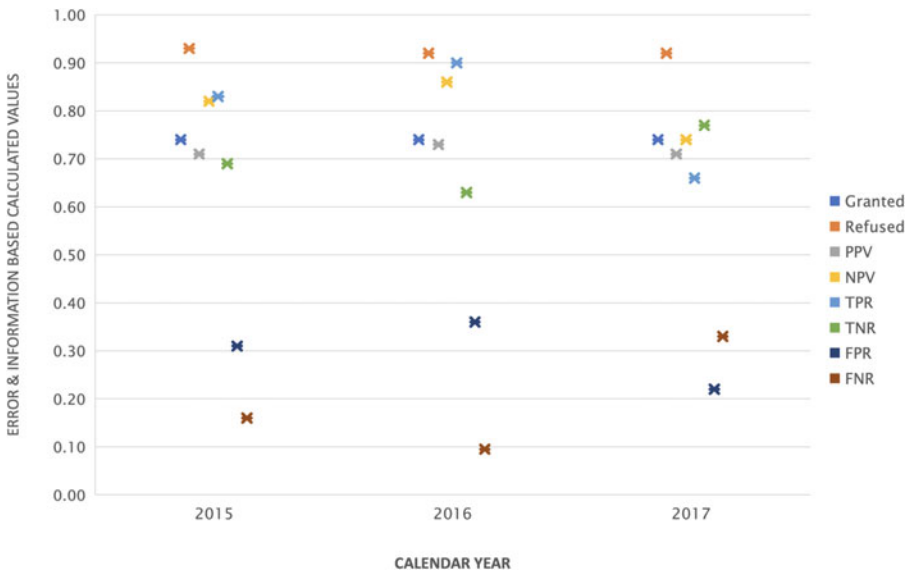


Figure 11. Comparison of probability distribution to the error-based and information-based values from *Bail-14*. PPV, positive predictive value; NPV, negative predictive value; TPR, true positive ratio; TNR, true negative ratio; FPR, false positive ratio; FNR, false negative ratio.

compliance given an outcome of Granted and the probability of non-compliance given an outcome of Granted – a corresponding pattern emerges (see Figure 12).

Discussion

This study synthesized real-world data using the approaches of B-LogR and tree-structured classification to demonstrate real-world outcomes. Fundamentally, this

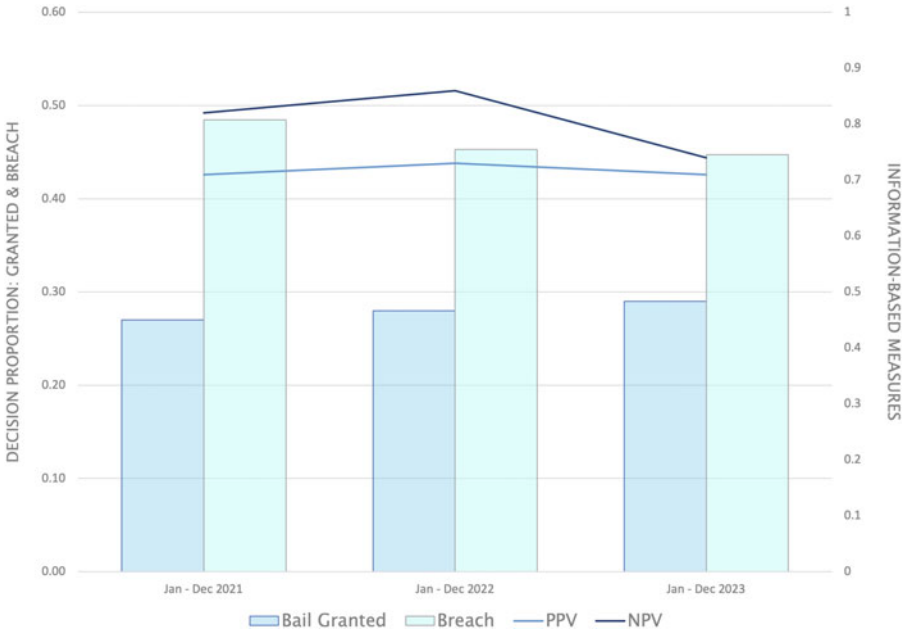


Figure 12. A three-year comparison of information-based measures positive predictive value (PPV) and negative predictive value (NPV) to bail granted and breach of bail.

was a process cognizant of balance and integrity principles – aimed at appeasement of the discourse around predictive modelling being used as decision-making means, where decisions could result in one’s liberty being wrongly denied (FP) or liberty being given incorrectly (FN). Values in this classification metric aim for the ideal; yet, in reality, these models could be considered dubious by some, as they challenge benchmarks on acceptable policy (Berk and Bleich 2013).

The literature on predictive modelling in crime-related domains assessing risk and recidivism reflects the results of this study (although noting wording variations within the literature does not detract from the premise of meaning). For example, a meta-analysis of predictor domains measuring recidivism found that criminal history was a prominent feature (Gendreau et al. 1996). Other scholarly work on AI and criminal justice, assessing recidivism and risk, reported predictor factors similar to those for offence seriousness, current and prior violence, prior convictions and failure to appear in court (Dupont et al. 2018). Notwithstanding, there could be a material reason why the “seriousness of offence(s)” in the T-sC was a prominent attribute, that is, the “gain criterion” having a preference for those attributes with a higher amount of values (Quinlan 1986).

Although there was no preconceived intention to compare the models used in this study, inferences were made from the EDA approach regarding the performance or strength of the models in terms of their predictive capability. B-LogR and the T-sC yielded some persuasive outcomes, notably with an overall accuracy in the 70th percentile. Notably, this level of accuracy is comparable to that

Table 14. Predictor relevance order based on Figure 7

Relevance	Predictors (L)	Predictors (R)
1	Seriousness of offences(s)	Seriousness of offences(s)
2	Show cause (yes/no)	Pro-criminal associations
3	Criminal history	Seriousness of offences(s)
4	Criminal history	History of violence
5	Bail non-compliance	Failure to appear/flight risk
6	History of violence	Show cause (yes/no)
7	Pro-criminal associations	–

Note: left-side tree (L) – criminal history node repeats although provides two different binary outcomes as expected; right-side tree (R) stopped at the sixth node.

of a predictive-based study on court decisions, which achieved 79% accuracy (Sourdin 2018). The TPR in all B-LogR models yielded a strong result, indicating a good predictive model, and the ROC values from *Model 51-50* and its sub-models offer some indication of its robustness. Comparatively, the T-sC values were lower, concluding that its predictive strength was secondary to B-LogR. A contrasting outcome was reported in research comparing logistic regression to decision tree models for predicting recidivism, where the decision tree approach outperformed the logistic regression approach (Wijenayake et al. 2018).

Statistical prominence was demonstrated in the T-sC model for both manipulated predictors, which is consistent with domain insight, where a defendant's criminal history and the seriousness of the offence are expected to be considered when determining bail. Other research conducted on decision-making in criminal justice also supported this contention (Lytle 2013). Analysing the data output from B-LogR and the T-sC yields results that are consistent with similar conclusions from the literature, namely, "seriousness of offence(s)" and, thereafter, "criminal history", which are consequential for predictive modelling of bail. Notably, the groundwork from this study builds upon similar overtures made by proponents in the literature, advocating for predictive modelling to be considered as a decision-making means in criminal matters.

Limitations

This study was retrospective, meaning the data were obtained from past cases and then analysed about the outcome of the already determined conviction or appeal. As will be suggested in future work, a different approach to retrospective analysis, such as real-time analysis, could be constructive. Along with the limitations of a retrospective study of this nature was the use of open-sourced case narratives. This meant that the quantity and quality of published judgments for bail hearings were a factor in the data numbers. First judges (or other administrative authorities) are selective on the cases they want published, and bail decisions appear to feature less than other decisions, such as sentencing. Second, some cases were restricted and not

available at the time of collection. Third, of the select published bail decisions, the key variables are not contained within all written judgments and could not be used. Therefore, quality and quantity negatively affect the total numbers and, consequently, this means a low number for training and testing. A more favourable scenario would be one where a greater number of published judgments and decisions accommodates larger datasets for training, testing and validation.

Collecting data from the case narratives was a laborious exercise – searching for the keywords and phrasing required each judgment to be read in its entirety, then re-read at different junctures to ensure the key variables were registered accurately. A limitation may be drawn on the data being specific to NSW and not any other, given that the states of Victoria and Queensland both apply the show cause test in their bail legislation.

Lastly, a small yet valid criticism can be made of the RapidMiner-Studio software. The tree classifier option did not definitively nominate an algorithm that draws on principles such as transparency and explainability, which serves as a criticism of predictive modelling more generally.

Conclusion and Scope for Future Work

Incontrovertibly, traditional decision-making for bail is based on human prediction. As such, in NSW and Australian courts collectively, an authorized justice is tasked with determining whether an accused is likely or unlikely to comply with bail conditions and if that risk can be mitigated or ameliorated by such conditions or whether it is safer to err on the side of caution and remand an accused in custody. Humans are not infallible, and mistakes can occur. There are well-documented cases where erroneous decisions have had tragic consequences.

Evidently, the predictive models used in this study also led to erroneous decisions. The models were somewhat rudimentary despite the presence of errors and study limitations. That being said, the performance outcomes demonstrated, prospectively, opportunities for further experimentation. An example of this could be in the development of a longitudinal study in real time: perhaps data collection and subsequent model testing for bail using the nine predictors might be more robust in the first instance, that is, to collect the data at such a time when bail is granted using a large cohort of accused persons; then, following up in time intervals, to determine if any breaches of the bail conditions were recorded, until such a time the respective bail matter has been finalized, yet all measured against the human-based decision made by a bail authority.

A key theme in this study was to advocate for an AI prototype and pilot study to be tested in NSW and/or other Australian criminal courts, similar to the related work. The results of these studies, based on their successes, would create grounds for policy change. Given the relative successes of this study and those in the literature, there is evidence in support of Australian courts considering the use of AI-driven predictive modelling to inform bail decisions.

Competing interests. No potential competing interests were reported by the authors.

References

- Alikhademi, K., E. Drobina, D. Prioleau, B. Richardson, D. Purves, and J. E. Gilbert. 2022. "A Review of Predictive Policing from the Perspective of Fairness." *Artificial Intelligence and Law* 30:1–17.
- Allsop, J. 2019. "Technology and the Future of the Courts." *University of Queensland Law Journal* 38(1):1–14.
- Alvo, Mayer. 2022. "Tools for Machine Learning." Pp. 277–328 in *Statistical Inference and Machine Learning for Big Data*. Cham: Springer.
- Attewell, Paul and David Monaghan. 2019. *Data Mining for the Social Sciences: An Introduction*. Oakland, CA: University of California Press.
- Audit Office of New South Wales. 2019. "Managing Growth in the NSW Prison Population." 24 May 2019, retrieved 2 July 2025 (<https://www.audit.nsw.gov.au/our-work/reports/managing-growth-in-the-nsw-prison-population>).
- Barabas, Chelsea, Madars Virza, Karthik Dinakar, Joichi Ito, and Jonathan Zittrain. 2018. "Interventions Over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment." *Proceedings of Machine Learning Research* 81:62–76.
- Berk, R. and J. Bleich. 2013. "Statistical Procedures for Forecasting Criminal Behavior: A Comparative Assessment." *Criminology and Public Policy* 12(3):513–44.
- Berk, R. and A. Elzarka. 2020. "Almost Politically Acceptable Criminal Justice Risk Assessment." *Criminology and Public Policy* 19(4):1231–57.
- Berk, R., S. Sorenson, and G. Barnes. 2016. "Forecasting Domestic Violence. A Machine Learning Approach to Help Inform Arraignment Decisions." *Journal of Empirical Legal Studies* 13(1):94–115.
- Bogart, Steve. 2023. "SankeyMATIC: Build a Sankey Diagram." Retrieved 2 July 2025 (<https://sankeymatic.com/build/>).
- Britt, Chester L. and David Weisburg. 2010. "Logistic Regression Models for Categorical Outcome Variables." Pp. 649–82 in *Handbook of Quantitative Criminology*, edited by Alex R. Piquero and David Weisburd. New York: Springer.
- Chan, Y. 2004. "Basic Statistics for Doctors." *Singapore Medical Journal* 45(4):149–53.
- Clipper, S. 2016. "Predicting Failure to Appear: A Comparison of Statistical Techniques." PhD dissertation, University of Texas, Dallas, TX.
- Colectica. 2024. "Colectica for Microsoft Excel." Retrieved 2 July 2025 (<https://www.colectica.com/software/colecticaforexcel/>).
- Dahiru, T. 2008. "P-Value: A True Test of Statistical Significance? A Cautionary Note." *Annals of Ibadan Postgraduate Medicine* 6(1):21–6.
- Dupont, Benoît, Yuan Y. Stevens, Hannes Westermann, and Michael Joyce. 2018. "Artificial Intelligence in the Context of Crime and Criminal Justice." SSRN, retrieved 2 July 2025 (<https://doi.org.10.2139/ssrn.4883392>).
- Farmer, C., I. Parsons, and M. Bagaric. 2018. "Sentencing Inconsistencies: A Case Study." *Australian Law Journal* 92:517–28.
- Gendreau, G., T. Little, and C. Goggin. 1996. "A Meta-Analysis of the Predictors of Adult Offender Recidivism: What Works." *Criminology* 34(4):575–608.
- Grove, W., D. Zald, B. Lebow, B. Snitz, and C. Nelson. 2000. "Clinical Versus Mechanical Prediction: A Meta-Analysis." *Psychological Assessment* 12(1):19–30.
- Hall, M., D. Calabro, T. Sourdin, A. Stranieri, and J. Zeleznikow. 2005. "Supporting Discretionary Decision-Making with Information Technology: A Case Study in the Criminal Sentencing Jurisdiction." *University of Ottawa Law and Technology Journal* 2:1–36.
- Helmus, L. and K. M. Babchishin. 2017. "Primer on Risk Assessment and the Statistics Used to Evaluate its Accuracy." *Criminal Justice and Behavior* 44(1):8–25.
- Jankovic, D. 2021. "A Simple Interpretation of Logistic Regression Coefficients." *Towards Data Science*, 15 September 2021, retrieved 2 July 2025 (<https://www.towardsdatascience.com/a-simple-interpretation-of-logistic-regression-coefficients>).
- Johnson, R. E. and D. Wichern. 2013. *Applied Multivariate Statistical Analysis*. Harlow: Pearson Education.
- Jung, J., C. Concannon, R. Shroff, S. Goel, and D. Goldstein. 2020. "Simple Rules to Guide Expert Classifications." *Journal of the Royal Statistics in Society Series A* 183(3):771–800.

- Kleinbaum, D. G., L. L. Kupper, A. Nizam, and E. S. Rosenberg. 2014. *Applied Regression Analysis and Other Multivariable Methods*, 5th edn. Boston, MA: Cengage Learning.
- Kotsiantis, S. 2013. "Decision Trees: A Recent Overview." *Artificial Intelligence Review* 39:261–83.
- Larner, A. J. 2021. *The 2x2 Matrix: Contingency, Confusion and the Metrics of Binary Classification*. Cham: Springer International Publishing.
- Lee, V. E., L. Liu, and R. Jin. 2014. "Decision Trees: Theory and Algorithms." Pp. 87–120 in *Data Classification: Algorithms and Applications*, edited by C. C. Aggarwal. New York: Chapman and Hall/CRC.
- Lotfi, I. and A. E. Bouhadi. 2022. "Artificial Intelligence Methods: Toward a New Decision-Making Tool." *Applied Artificial Intelligence* 36(1):101–6.
- Lum, K., C. Boudin, and M. Price. 2020. "The Impact of Overbooking on a Pre-Trial Risk Assessment Tool." *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 482–91, retrieved 2 July 2025 (<https://doi.org.10.1145/3351095.3372846>).
- Lussier, P., N. Deslauriers-Varin, J. Collin-Santerre, and R. Bélanger. 2019. "Using Decision Tree Algorithms to Screen Individuals at Risk of Entry into Sexual Recidivism." *Journal of Criminal Justice* 63:12–24.
- Lytle, Daniel J. 2013. "Decision Making in Criminal Justice Revisited: Toward a General Theory of Criminal Justice." PhD dissertation, University of Cincinnati, OH.
- Martino, J. 2019. *Hands-On Machine Learning with Microsoft Excel 2019: Build Complete Data Analysis Flows, from Data Collection to Visualization*. Birmingham: Packt Publishing Ltd.
- McCue, C. 2014. *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis*. Oxford: Butterworth Heinemann.
- Morgan, A. 2018. "How Much Does Prison Really Cost? Comparing the Costs of Imprisonment with Community Corrections." Australian Institute of Criminology, Research Report 05, retrieved 2 July 2025 (https://www.aic.gov.au/sites/default/files/2020-05/rr_05_240418_2.pdf).
- Nasridinov, Aziz, Sun-Young Ihm, and Young-Ho Park. 2013. "A Decision Tree-Based Classification Model for Crime Prediction." Pp. 531–8 in *Information Technology Convergence: Security, Robotics, Automations and Communication*, edited by James J. Park, Leonard Barolli, Fatos Xhafa, and Hwa-Young Jeong. New York: Springer.
- Ngo, F., R. Govindu, and A. Agarwal. 2015. "Assessing the Predictive Utility of Logistic Regression, Classification and Regression Tree, Chi-Squared Automatic Interaction Detection, and Neural Network Models in Predicting Inmate Misconduct." *American Journal of Criminal Justice* 40(1):47–74.
- NSW Bureau of Crime Statistics and Research. 2023. "Criminal Court Statistics." (<https://bocsr.nsw.gov.au/statistics-dashboards/court-and-sentencing.html>).
- Organisation for Economic Co-operation and Development. 2025. "Recommendation of the Council on Artificial Intelligence." Retrieved 30 June 2025 (https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449?_ga=2.124697866.1898590258.1621541868-1852956558.1620326402).
- Pennsylvania State University. 2018. "Applied Regression Analysis." Department of Statistics Online Programs. Retrieved 2 July 2025 (<https://online.stat.psu.edu/stat462/node/207/>).
- Pennsylvania State University. 2024. "Probability Distributions." Retrieved 2 July 2025 (<https://online.stat.psu.edu/stat500/lesson/3/3.1>).
- Pinder, J. P. 2020. *Introduction to Business Analytics Using Simulation*, 2nd edn. London: Academic Press.
- Productivity Commission. 2021. "Australia's Prison Dilemma: Commission Research Paper." Australian Government Productivity Commission, 29 October 2021, retrieved 2 July 2025 (<https://www.pc.gov.au/research/completed/prison-dilemma>).
- Quinlan, J. R. 1986. "Induction of Decision Trees." *Machine Learning* 1:81–106.
- Ramakrishnan, Naren. 2009. "C4.5." Pp. 1–20 in *The Top Ten Algorithms in Data Mining*, edited by Xindong Wu and Vipin Kumar. Boca Raton, FL: CRC Press.
- RapidMiner. 2024. "RapidMiner-Studio, Version 10.2." Retrieved 2 July 2025 (www.rapidminer.com).
- Rutkowski, Leszek, Maciej Jaworski, and Piotr Duda. 2020. "Decision Trees in the Data Stream Mining." Pp. 37–50 in *Stream Data Mining: Algorithms and their Probabilistic Properties*. Springer International Publishing.
- Shaikh, R., T. Sahu, and V. Anand. 2020. "Predicting Outcomes of Legal Cases Based on Legal Factors Using Classifiers." *Procedia Computer Science* 167:2393–402.
- Shmueli, G. 2010. "To Explain or to Predict?" *Statistical Science* 25(3):289–310.

- Singh, T., Y. Jain, and V. Kumar. 2017. "Predicting Parole Hearing Result Using Machine Learning." 2017 *International Conference on Emerging Trends in Computing and Communication Technologies (ICETCCT)*, Dehradun, India. Retrieved 2 July 2025 (<https://doi.org/10.1109/ICETCCT.2017.8280342>).
- Skeem, J. and C. Lowenkamp. 2020. "Using Algorithms to Address Trade-Offs Inherent in Predicting Recidivism." *Behavioral Sciences and the Law* 38(3):259–78.
- Sourdin, Tania. 2018. "Judge v Robot: Artificial Intelligence and Judicial Decision-Making." *UNSW Law Journal* 41(4):1114–33.
- Stobbs, N., D. Hunter, and M. Bagaric. 2017. "Can Sentencing be Enhanced by the Use of Artificial Intelligence?" *Criminal Law Journal* 41(5):261–77.
- Waltl, B. and R. Vogl. 2018. "Explainable Artificial Intelligence – The New Frontier in Legal Informatics." *Jusletter IT* 4:1–10, retrieved 2 July 2025 (<https://www.matthes.in.tum.de/file/13tkeaid0rhkz/Sebis-Public-Website/-/Explainable-Artificial-Intelligence-the-New-Frontier-in-Legal-Informatics/Wa18a.pdf>).
- Wijenayake, S., T. Graham, and P. Christen. 2018. "A Decision Tree Approach to Predicting Recidivism in Domestic Violence." Pp. 3–15 in *Trends and Applications in Knowledge Discovery and Data Mining. PAKDD 2018. Lecture Notes in Computer Science*, Vol. 11154, edited by M. Ganji, L. Rashidi, B. C. M. Fung, and C. Wang. Cham: Springer.
- Wilkinson, J. D., M. A. Mamas, and E. Kontopantelis. 2022. "Logistic Regression Frequently Outperformed Propensity Score Methods, Especially for Large Datasets: A Simulation Study." *Journal of Clinical Epidemiology*. 152:176–84.
- Wykstra, S. 2020. "Governance, Risk and Artificial Intelligence." *AIMag* 41(1):61–9.
- Zaidi, A. and A. S. M. Al Luhayb. 2023. "Two Statistical Approaches to Justify the Use of the Logistic Function in Binary Logistic Regression." *Mathematical Problems in Engineering* 2023:525675.
- Zaiontz, C. 2024. "Real Statistics Resources Pack." *Real Statistics Using Excel*, retrieved 3 July 2025 (<https://www.real-statistics.com/free-download/real-statistics-resource-pack>).
- Zeng, J., B. Ustun, and C. Rudin. 2017. "Interpretable Classification Models for Recidivism Prediction." *Journal of the Royal Statistical Society Series A* 180(3):689–722.
- Zettler, H. R. and R. G. Morris. 2015. "An Exploratory Assessment of Race and Gender-Specific Predictors of Failure to Appear in Court Among Defendants Released Via a Pretrial Services Agency." *Criminal Justice Review* 40(4):417–30.

Cases Cited

- DPP (NSW) v. Zaiter [2016] NSWCCA 247.
 JM v. R [2015] NSWSC 978.
 M v. R [2015] NSWSC 138.
 R v. Peter Tsallas [2017] NSWSC 64.

Legislation

- Bail Act 2013 (NSW). Retrieved 2 July 2025 (<https://legislation.nsw.gov.au/view/html/inforce/current/act-2013-026>).
- Bail Amendment Bill 2014 (NSW), Second Reading. Retrieved 2 July 2025 (<https://www.parliament.nsw.gov.au/Hansard/Pages/HansardResult.aspx#/docid/HANSARD-1323879322-57674:~:text=Debate%20resumed%20from%20an%20earlier,bail%20legislation%20in%20this%20State>).
- Bail Amendment Bill 2015 (NSW), Second Reading. Retrieved 2 July 2025 (<https://www.parliament.nsw.gov.au/Hansard/Pages/HansardResult.aspx#/docid/HANSARD-1323879322-67864>).

Translated Abstracts

Abstracto

El mundo inteligible de las máquinas y el modelado predictivo es un fenómeno omnipresente y casi ineludible. Se trata de una evolución en la que la inteligencia humana se ve apoyada, complementada o incluso superada por la Inteligencia Artificial (IA). Las decisiones que antes tomaban los humanos ahora las toman las máquinas, que aprenden a un ritmo más rápido y preciso mediante cálculos algorítmicos. La academia jurisprudencial se ha propuesto argumentar la propuesta de la IA y su papel como mecanismo de toma de decisiones en las jurisdicciones penales australianas. Este artículo explora esta propuesta mediante el modelado predictivo de 101 decisiones sobre fianzas tomadas en tres tribunales penales del estado de Nueva Gales del Sur (NSW), Australia. Como muestra, el rendimiento estadístico y la precisión de los modelos, basados en nueve variables predictoras, demostraron ser eficaces. El modelo de regresión logística, más preciso, alcanzó una precisión del 78 % y un valor de rendimiento de 0,845 (AUC), mientras que el modelo clasificador alcanzó una precisión del 72,5 % y un valor de rendimiento de 0,702 (AUC). Estos resultados proporcionan las bases para las decisiones de fianza generadas por IA que se están probando en la jurisdicción de Nueva Gales del Sur y posiblemente en otras dentro de Australia.

Palabras clave: fianza; toma de decisiones; tribunales penales; aprendizaje automático; inteligencia artificial; regresión logística binaria; clasificador estructurado en árbol

Abstrait

Le monde intelligible des machines et de la modélisation prédictive est un phénomène omniprésent et quasi incontournable. Il s'agit d'une évolution où l'intelligence humaine est soutenue, complétée ou supplantée par l'intelligence artificielle (IA). Les décisions autrefois prises par les humains sont désormais prises par des machines, qui apprennent plus rapidement et avec plus de précision grâce à des calculs algorithmiques. Des universitaires ont entrepris de défendre la thèse de l'IA et son rôle comme mécanisme décisionnel dans les juridictions pénales australiennes. Cet article explore cette thèse à travers la modélisation prédictive de 101 décisions de mise en liberté sous caution rendues par trois tribunaux pénaux de l'État de Nouvelle-Galles du Sud (NSW), en Australie. À titre indicatif, les performances statistiques et la précision des modèles, basés sur neuf variables prédictives, se sont avérées efficaces. Le modèle de régression logistique, plus précis, a atteint une précision de 78 % et une valeur de performance de 0,845 (AUC), tandis que le modèle de classification a atteint une précision de 72,5 % et une valeur de performance de 0,702 (AUC). Ces résultats constituent la base des décisions de mise en liberté sous caution générées par l'IA, actuellement testées dans la juridiction de la Nouvelle-Galles du Sud et peut-être dans d'autres juridictions en Australie.

Mots clés: cautionnement; prise de décision; tribunaux pénaux; apprentissage automatique; intelligence artificielle; régression logistique binaire; classificateur arborescent

摘要

机器和预测模型的可理解世界是一个无处不在且几乎不可避免的现象。这是一种进化，人类智能正在被人工智能 (AI) 支持、补充或取代。曾经由人类做出的决策现在由机器做出，通过算法计算，机器能够以更快、更准确的速度学习。法学界已开始探讨人工智能及其作为澳大利亚刑事司法决策机制的作用。本文通过对澳大利亚新南威尔士州 (NSW) 三个刑事法庭的 101 个保释判决进行预测建模，探讨了这一命题。结果表明，基于九个预测变量的模型的统计性能和准确性得到了验证。更精确的逻辑回归模型达到了 78% 的准确率和 0.845 (AUC) 的性能值，而分类器模型达到了 72.5% 的准确率和 0.702 (AUC) 的性能值。这些结果为在新南威尔士州乃至澳大利亚其他地区试行的人工智能保释决定奠定了基础。

关键词: 保释；决策；刑事法庭；机器学习；人工智能；二元逻辑回归；树状分类器

ملخص

يُعدّ عالم الآلات والنمذجة التنبؤية، المفهوم، ظاهرةً حاضرةً في كل مكان، ويكاد يكون حتميًا. إنه تطورٌ يُدعم فيه الذكاء الاصطناعي الذكاء البشري، أو يُكمّله، أو يُحلّ محله. فالقرارات التي كان يتخذها البشر سابقًا، تتخذ الآن بواسطة الآلات، حيث تتعلم بمعدل أسرع وأكثر دقة من خلال الحسابات الخوارزمية. وقد بادرت الأوساط الأكاديمية الفقهية إلى مناقشة فرضية الذكاء الاصطناعي ودوره كآلية لصنع القرار في المحاكم الجنائية الأسترالية. تستكشف هذه الورقة هذه الفرضية من خلال النمذجة التنبؤية لـ 101 قرار إخلاء سبيل بكفالة صادرة عن ثلاث محاكم جنائية في ولاية نيو ساوث ويلز، أستراليا. ومن الجدير بالذكر أن الأداء الإحصائي ودقة النماذج، المستندة إلى تسعة متغيرات تنبؤية، أثبتت فعاليتها. حقق نموذج الانحدار اللوجستي الأكثر دقةً دقةً بنسبة 78% وقيمة أداء 0.845 (AUC)، بينما حقق نموذج التصنيف دقةً بنسبة 72.5% وقيمة أداء 0.702 (AUC). توفر هذه النتائج الأساس لقرارات الكفالة التي يتم إنشاؤها بواسطة الذكاء الاصطناعي والتي يتم تجربتها في ولاية نيو ساوث ويلز وربما مناطق أخرى داخل أستراليا.

الكلمات المفتاحية: الكفالة، اتخاذ القرار، المحاكم الجنائية، التعلم الآلي، الذكاء الاصطناعي، الانحدار اللوجستي، التصنيف، المصنف الهيكلي، الشجر، الذكاء الاصطناعي.

Appendix

Model Predictor Information Table

Predictors and descriptions	History of violence		Coding / Value
	Prior charges, convictions regarding violence Recorded information from custody status (matters of violence while in police or corrections custody)	No [0] Yes [1]	
	Prior charge or conviction while on bail (in NSW or another jurisdiction)		
	Defendant has breached bail conditions on one or more occasions, even if it did not result in a court hearing.	No [0] Yes [1]	
	History of non-compliance with court-issued orders		
	Defendant has demonstrated non-compliance with orders for community-based supervision (e.g., parole).	No [0] Yes [1]	
	Pro-criminal associations		
	• Defendant is a member or linked to organised criminal network, gang affiliation. • Current offences were committed in company of co-accused. • Defendant has history of offending in company of other convicted offenders.	No [0] Yes [1]	
	Danger that may be posed to the public/victim(s) or other persons known to them		
	• Defendant has previously demonstrated threats, did or had attempted to contact victim or witnesses. • Defendant has currently or previously demonstrated disregard for public safety.	No [0] Yes [1]	
	Flight risk/failure to attend		
	• Flee a jurisdiction; ability to obtain falsified identification, documents, passports. • Defendant has prior occasions of failing to appear at court.	No [0] Yes [1]	

Brett Anthony Hansard completed his research at University of Technology Sydney (UTS) in 2025 on mitigating the ethical considerations of artificial intelligence through visual analytics. This research was supported by the Australian Government Research Training Programme. Brett's undergraduate and postgraduate studies were in the social science disciplines of criminology and sociology. Brett has previously been employed at the Justice Department in NSW.

Jianlong Zhou is an Associate Professor in the School of Computer Science, Faculty of Engineering and Information Technology, UTS, leading the Human Centred AI research laboratory. He has extensive research experience in various fields including AI, visual analytics, virtual reality/augmented reality, and human-computer interaction in different universities and research institutes in the United States, Germany and Australia. Prior to joining UTS, Jianlong was a senior research scientist at various Australian government agencies.