

Case Commentary: Artificial Intelligence in Malaysian Courts: PP v Denis P Modili

Dennis W K Khong

Centre for Law and Technology, Faculty of Law,
Multimedia University, Malaysia

wkkhong@mmu.edu.my

ORCID iD: 0000-0002-9730-0944

(Corresponding author)

Chiung Ching Ho

Sunway University, Malaysia

peterh@sunway.edu.my

ORCID iD: 0000-0001-5098-8131

ABSTRACT

This case commentary examines a seminal judgment on the application of artificial intelligence in criminal sentencing in Malaysia: *PP v Denis P Modili*. It hypothesises what the basic components of this AI system may consist of and discusses some technical points based on the disclosure made about the AI system in the judgment. Two types of machine learning algorithms, namely linear regression and logistic regression, are introduced. Possible machine learning features and their treatment, and what the 'recommended percentage' as described in the judgment could be are investigated.

Keywords: Sentencing recommendation system, Artificial intelligence, Machine learning

Received: 26 May 2022, **Accepted:** 30 May 2022, **Published:**

1. Introduction

In January 2019, during the opening of the legal year, the then Chief Justice of the Federal Court of Malaysia, Tun Richard Malanjum, disclosed that plans were afoot to provide for judges and judicial officers with an artificial intelligence (AI) system to minimise disparity in

criminal sentences.¹ One year later, the then Chief Judge of Sabah and Sarawak, Tan Sri Datuk Seri Panglima David Wong Dak Wah announced that the courts in Sabah and Sarawak will pioneer a trial for the AI for Sentencing application, which at that moment, was confined to sentencing for drug possession offences under section 12(2) of the Dangerous Drugs Act 1952 and rape offences under section 376 of the Penal Code.² It was further reported that one of the reasons for initially choosing these two sections is that ‘the dataset for those two offences is the richest dataset that they have’.³ Eventually, the AI sentencing guidelines (AIGS) system is known as the Artificial Intelligence in Court Sentencing (AiCOS)⁴ system and was developed by a state government-owned company, Sarawak Information Systems Sdn Bhd.⁵ This pioneering effort is in line with a global movement to use artificial intelligence in court processes and criminal sentencing.⁶

The opportunity to use the AI sentencing system arose in Sabah in a drug possession case, in which the sitting magistrate employed the AI sentencing system to recommend the term of imprisonment of an accused. Not content with this approach, the accused made an appeal for which a written judgment was provided by the magistrate outlining her decision for using the AI sentencing system. However, when the appeal went to the High Court, the whereabouts of the appellant were untraceable and consequently, no legal argument was made. Noting that the twelve-month imprisonment sentence has already been completed, the High Court reduced the sentence to six months but provided no written justification for it.⁷

This case commentary examines this seminal judgment on the application of artificial intelligence in criminal sentencing in Malaysia: *PP v Denis P Modili*,⁸ hypothesises what the basic components of this AI system may consist of, and discusses some technical points based on the disclosure made about the AI system in the judgment.

¹ Richard Malanjum, ‘Speech by Tan Sri Datuk Seri Panglima Richard Malanjum, Chief Justice of the Federal Court of Malaysia at the Opening of the Legal Year 2019’ [2019] *Malayan Law Journal* i.

² Neil Brian Joseph, ‘KK Courts Set to Be First to Use Artificial Intelligence’ *Borneo Post Online* (31 January 2020).

³ Claire Lim and Rachel Gong, ‘Artificial Intelligence in the Courts: AI Sentencing in Sabah and Sarawak’ (2020) 2 <<http://www.krinstitute.org/assets/contentMS/img/template/editor/210206 AI in the Courts v4.pdf>>.

⁴ Abang Iskandar Bin Abang Hashim, ‘Speech by the Right Honourable Tan Sri Dato’ Abang Iskandar Bin Abang Hashim, Chief Judge of Sabah and Sarawak, at the Opening of the Legal Year Sabah and Sarawak 2022’ (2022) <https://judiciary.kehakiman.gov.my/portals/web/home/article_view/0/1803/>.

⁵ Bernama, ‘Sabah, Sarawak Courts to Enhance Digitalisation Effort: Chief Judge’ *The Sun Daily* (21 January 2022) <<https://www.thesundaily.my/local/sabah-s-wak-courts-to-enhance-digitalisation-effort-chief-judge-FJ8784124>>; Pathma Subramaniam, ‘Programmed Prejudice’ *Digital Edge: The Edge Weekly* (4 October 2021) <<https://www.theedgemarkets.com/article/cover-story-programmed-prejudice>>.

⁶ Jesper Ryberg and Julian V Roberts (eds), *Sentencing and Artificial Intelligence* (Oxford University Press 2021).

⁷ *Denis P Modili v Public Prosecutor* (HC, Kota Kinabalu, 15 September 2020); Cheng Leong Foong, ‘The Use of Court’s Artificial Intelligence (AI) System in Criminal Proceedings’ (2021) <<http://foongchingleong.com/2021/06/>>.

⁸ [2020] 2 Sessions & Magistrates’ Cases 381 (Magistrate’s Court).

2. Facts and Legal Issues

The accused was charged and convicted by the Magistrate for possession of 0.01 gram of methamphetamine under section 12(2) of the Dangerous Drugs Act 1952. Punishment for an offence under that section is 'a fine not exceeding one hundred thousand ringgit or to imprisonment for a term not exceeding five years or to both'.⁹ In addition to the charge under section 12(2), the accused was also concurrently charged under section 15(1)(a) of the Dangerous Drugs Act 1952 as a user of a dangerous drug. At the Magistrate's Court, the accused pleaded guilty to the charges, and after using the AI sentencing system, the magistrate sentenced the accused to twelve months imprisonment from the date of arrest for the offence under section 12, to run concurrently with the sentence under section 15. Thereafter, the accused filed a notice of appeal against the sentence.

The *pro bono* counsel for the accused objected to the court using the AI sentencing system. It was argued that the use of the AI sentencing system is against Article 5(1) and Article 8(1) of the Federal Constitution. Article 5(1) provides for the protection of the life and personal liberty of a person, while Article 8(1) guarantees equality before the law and equal protection of the law. The Magistrate's answer to the arguments raised by the defence counsel was that since constitutional issues are not within the competent jurisdiction of a Magistrate's court, the issues can only be determined by the High Court on appeal.

Proceeding to use the AI sentencing system, the Magistrate entered the required values into the AI system and obtained a recommendation of an imprisonment term of ten months with a 'probability of 54.31%'.¹⁰ The court mentioned three values which were required, namely 'the weight of the drugs, the age and employment record of the accused',¹¹ but she did not elaborate on whether any other value was used.

The procedure adopted by the court with regards to the usage of the AI sentencing system is to first read the recommendation from the AI system to the accused, then explain the functioning of the AI system, and finally offer an opportunity to the accused to change his plea if he so wishes.¹² This procedure is adopted by the court to ensure 'a fair and just hearing' with regard to the usage of the AI sentencing system.¹³ Interestingly, one question remains on how to situate this procedure to section 173(b) of the Malaysian Criminal Procedure Code and the twelve steps laid down in *Aung Min Aung v PP*.¹⁴ One possible interpretation is that this AI procedure is considered as part of the plea in mitigation.

In the present case, after the accused affirmed his plea of guilt, and notwithstanding that the AI sentencing system recommended imprisonment of ten months, the court sentenced

⁹ Dangerous Drugs Act 1952 (Act 234) s 12(3).

¹⁰ *Denis P Modili* (n 2) para 20.

¹¹ *Denis P Modili* (n 2) para 11.

¹² *Denis P Modili* (n 2) para 21.

¹³ *Denis P Modili* (n 2) para 22.

¹⁴ *Maung Min Aung v Public Prosecutor* [2001] 5 Malayan Law Journal 140 (HC).

the accused to imprisonment of twelve months, in order to 'not only give a clear message to the offenders out there but the-would-be offender as well'.¹⁵

3. Constructing an AI Sentencing System

Four points of observation regarding the design of the artificial intelligence system can be gleaned from the published report:

- (i) Data collected were based on the courts' database from 2014 to 2019,¹⁶
- (ii) At least three features are captured: weight of the drugs, the age and employment record of the accused,¹⁷
- (iii) The recommendations are either a sentence of fine or imprisonment,¹⁸
- (iv) A recommended percentage is given 'in order to assist the presiding judge to decide in applying the correct sentencing principles as what was decided by past precedents'.¹⁹ It appears that this 'recommended percentage' is given for both fines and imprisonments, as evidenced from the statement in the judgement: 'The recommended percentage derived from the AI is imprisonment of ten months which is based on a higher estimated probability of 54.31%'.²⁰

3.1 Machine Learning Algorithms

Most of the current crops of artificial intelligence systems use a statistical-based technique known as machine learning. The term 'machine learning' is an umbrella term for a class of many different algorithms which can be used to perform artificial intelligence functions after being 'trained' with an appropriate dataset known as 'training data'. Usually, these datasets are formatted in rows and columns, with each row representing an individual case, and columns representing values corresponding to various characteristics or features of that case, e.g. weight of the drug, age and employment status of the accused.²¹ These features in the columns are also known as input variables. Also captured in the columns are the dependent or target variables, which in the case of an AI sentencing system are the amount of fines and period of imprisonment.

The objective of machine learning is to use the training dataset to create a system which will predict or suggest the best dependent values for a new set of input values. The machine

¹⁵ *Denis P Modili* (n 2) para 24.

¹⁶ *Denis P Modili* (n 2) para 10.

¹⁷ *Denis P Modili* (n 2) para 11.

¹⁸ *Denis P Modili* (n 2) para 11.

¹⁹ *Denis P Modili* (n 2) para 12.

²⁰ *Denis P Modili* (n 2) para 20.

²¹ Dennis WK Khong and Wan-Ju Yeh, 'The Problem of Copyright Protection for Machine Learning Databases: A Comparative Study' (2021) 10 NTUT Journal of Intellectual Property Law and Management 86.

learning system which does this is called a model, and the model performs its task using a pre-selected algorithm. As there are hundreds of possible algorithms to use, it is the job of the AI developer to identify a suitable algorithm and to transform the variables into a suitable form to create the best model which can perform its job. In doing so, various trade-offs have to be made, such as preventing both over-fitting (high variance) and under-fitting (high bias) of a model to its training data. An over-fitted model may predict very well the training data but not generalise well with new input values, while an under-fitted model does not discriminate enough different input values to generate the predicted value. Examples of other trade-offs that an AI developer needs to take into account include the number of input variables needed as well as the computational processing power required to create and utilise the model.

3.2 Linear Regression Models

Linear regression is a statistical technique to predict dependent values which are numerical in nature. A multiple linear regression takes more than one input value to predict one output value. For example, a sentencing guideline system which predicts an imprisonment term based on several input values can be developed using multiple linear regression.

A more sophisticated type of linear regression model is a multivariate regression where more than one predicted value are obtained²². Hence, in a system where both fines and imprisonment are ordered by the courts, a multivariate regression model would be a suitable model to make such predictions.

Hypothetically, three models can be developed using linear regression: (i) a multiple regression with amount of fines as the predicted variable, (ii) a multiple regression with imprisonment term as the predicted variable, and (iii) a multivariate regression with both fines and imprisonment term as the predicted variables. The first model is built using training data where only fines were imposed, the second model using training data where only imprisonment was ordered, while the third model can be developed using all the available training data by giving a zero value if no fines or no imprisonment was ordered. The advantage of a three-model setup is that the judge using such an AI sentencing system can choose whether he wishes to impose only fines, only imprisonment, or a combination of fines and imprisonment.

A note can be made about the nature of the values for fines and imprisonment. Fines are typically imposed in hundred and thousands; henceforth, there is no smooth transition of values on a linear scale. Similarly, apart from the occasionally nominal imprisonment of a day or two, the terms of imprisonment are typically ordered in weeks, months or years and not in days. Thus, there is a need to convert the training data into days or weeks, and likewise convert the predicted values back to months and years, and round the values

²² John Fox and Sanford Weisberg, 'Multivariate Linear Models in R', *An R companion to applied regression* (3rd edn, SAGE Publications 2018) <<https://socialsciences.mcmaster.ca/jfox/Books/Companion/appendices/Appendix-Multivariate-Linear-Models.pdf>>.

accordingly. In the same way, the predicted fines may need to be rounded accordingly as typically ordered by the courts. From a statistical perspective, rounding of the predicted values is a form of distortion which makes the machine learning system less reliable.

3.3 Classification

A different question may be asked about how to develop a system to recommend whether the sentences should be fines only, imprisonment only or a mix of fines and imprisonment. This kind of machine learning function falls within what is known as classification. A typical example of classification is to identify whether a given email text is spam or not spam, or given a picture of a number, what number that picture shows.

To develop a sentencing classifier, the dependent variable in the training dataset will need to be in the form of a factor variable containing one of three possible choices: fines only (which can be represented as a '0'), imprisonment only (represented as a '1') and both fines and imprisonment (represented as a '2').

Typical machine learning algorithms which can be used to build a classifier are logistic regression, support vector machine and decision tree.²³ In practice, which algorithm performs better than the others depends on the nature of the input variables and the number of cases (rows).

A multiclass classifier is a classification technique where the dependent variable can have a choice of values.²⁴ Some machine learning algorithms can perform multiclass classification natively, while others such as logistic regression and support vector machines are only capable of handling dependent variables with two values, i.e. '0' and '1'. In such a situation, a strategy called one-versus-the-rest or one-versus-all is used by creating three separate binary classifiers: (i) fines only versus the rest, (ii) imprisonment only versus the rest, and (iii) fines and imprisonment versus the rest. In deployment, all three classification models are run using the same input data, and the model which gives the highest predicted value in a probability form is selected as the recommendation for sentences.

Finally, it is possible to combine the developed classification model with the regression models in a two-step process to create a seamless recommendation system on whether to impose a sentence of only a fine, only imprisonment or a combination of both, as well as to recommend the amount for each.

²³ Aurélien Géron, *Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow* (2nd edn, O'Reilly 2019) chs 3–6.

²⁴ Géron (n 23) 100.

4. Discussion

Based on the four observations from the *Denis P Modili* report above, two further points of discussion can be made: first, on the input variables that should be captured to develop the training data, and secondly, on what the ‘recommended percentage’ could be.

4.1 Input Variables

Although it was mentioned that the ‘parameters’ that are being captured by the AI sentencing system are ‘the weight of the drugs, the age and employment record of the accused’,²⁵ there are possibly other input variables that can be captured to better reflect the dependent variables in the data. Note that in machine learning terminology, a ‘parameter’ is not the input data in the training data or the data entered by a user, but is the weight or coefficient used to multiply with an input variable. Parameters are what turn a generic algorithm into a specific model that can be deployed in operation.

Since different drugs have different dosages to be of any effect, it is not likely that all types of drugs will be treated in the same way in sentencing. Hence, one input variable will necessarily have to be included is the type of drugs. Such variables are known as factor variables or categorical variables. Another factor variable that can be captured in the training data is the gender of the accused.

American Legal Realist, Jerome N. Frank, introduced the concepts of ‘rule skepticism’ and ‘fact skepticism’.²⁶ Rule skepticism focuses chiefly on the behaviour of appeal judges who primarily decide on questions of law, and concludes that judges’ decisions may be influenced by factors outside of the language of the law. This is demonstratively clear in the case of justices of the United States Supreme Court where nominations are made by the sitting president of the United States and partly based on the likely political leaning of the nominees when deciding cases.²⁷

Fact skeptacists, as Jerome Franks associates himself with, argue that ‘trial judges or juries, also human, may have prejudices—often unconscious, unknown even to themselves—for or against some of the witnesses, or the parties to the suit, or the lawyers’²⁸ In conclusion, American Legal Realists believe that regardless of whether judges are deciding cases in a trial or on appeal, the idiosyncrasies and personalities of judges may play a role in their decision-making. Hence, in developing a training dataset of court decisions, it is necessary to capture the name of the magistrate or judge as a factor variable. Likely, the level of the court, namely, Magistrate’s Court, Sessions Court, High Court, Court of Appeal and Federal Court, may also be useful as an input variable in the training data.

²⁵ *Denis P Modili* (n 2) para 11.

²⁶ Jerome Frank, ‘Legal Thinking in Three Dimensions’ (1949) 1 *Syracuse Law Review* 9 <<http://hdl.handle.net/20.500.13051/3550>>.

²⁷ This, however, is not meant to suggest that the political leanings of magistrates and judges in Malaysia are as prominently apparent as in the United States.

²⁸ Frank (n 26) 12.

Whether these factor variables turn up to be significant or otherwise can be decided using appropriate statistical tests. If they do not have a strong explanatory value to the criminal sentences, they may be dropped in the model at the model-tuning stage.

Another class of input variables that would be very useful to be included in the training dataset is mitigating and aggravating factors. Many of the actual sentences given by magistrates would have taken into account mitigating and aggravating factors, although it is not clear in the courts' reports whether these factors are written up and accepted. Since there is no way of separating a sentence without mitigating and aggravating factors, it is inevitable that the model must include possible mitigating and aggravating factors so that the model correctly takes into account the influence of these factors in the final sentences. For example, one important aggravating factor to include in such a model is recidivism, either captured as a binary 'yes' or 'no' data or by the number of previous convictions for the same type of crime or the number of previous convictions under the Dangerous Drugs Act 1952.

Algorithmic or ethical bias is a situation where a model in deployment reproduces undesirable or illegal biases found in the training data, e.g. race of the accused, whereas in an ideal world, the court should be colour-blind and that the race of the accused is irrelevant to the sentence. From a statistical point of view, it is possible to detect the presence of ethical bias by including the bias as an input variable while training the machine learning model.

To have a fairer and potentially less biased AI sentencing system, one approach to debiasing such a model is to drop or mask the ethical bias input variable at the stage of deployment.²⁹ This approach relies on the idea that the parameters for the biased input variables will be able to absorb the bias caused by those variables, and excluding them in the deployed model will lead to a less biased prediction. For example, since a fair sentence should not be determined by the identity of the magistrate or judge and the level of the court, the input variables representing these should only be included in the model at the training stage, but excluded from the model when deployed.

It was reported that certain potentially biased variables, such as the race of the accused, were 'remove[d] ... from the algorithm as it was not a significant factor in the sentencing process'.³⁰ However, the Khazanah Research Institute's report made no mention of whether these bias variables were dropped after their parameters were estimated, or before the model was developed. The risk of not including 'race' in the model at the training stage is that should 'race' be a significant but latent factor in the actual sentencing of prior cases, then a proxy variable may emerge which captures the parameter of this excluded 'race' variable. For example, if education level is weakly correlated to race, then if race is excluded as a variable in the training data, the influence of race in the sentences will be reflected indirectly in education level.

²⁹ Soheil Ghili, Ehsan Kazemi and Amin Karbasi, 'Eliminating Latent Discrimination: Train Then Mask', *Proceedings of the AAAI Conference on Artificial Intelligence* (2019) <<https://doi.org/10.1609/aaai.v33i01.33013672>>.

³⁰ Lim and Gong (n 3).

It is also unfortunate that mitigating factors seem not to have been included in the training data.³¹ The fact that the deployed AI sentencing system does not allow magistrates and judges to input mitigating factors causes it to be less useful than otherwise. It was thus reported that as of 29 May 2020, judges have departed from the recommended sentences in 67% of the cases, one of which is the case of *PP v Denis P Modili*.

4.2 Recommended Percentage

In the case report, it was mentioned that '[t]he recommended percentage derived from the AI is imprisonment of ten months which is based on a higher estimated probability of 54.31%'. However, not much information can be derived from this passage as to what exactly this 'recommended percentage' represents from a statistical point of view.

In logistic regression, for every set of input values, there will be a predicted value, which is expressed as a probability that the outcome is true. Thus, when the predicted probability from a logistic regression predicting imprisonment (dependent variable = 1) or no imprisonment (dependent variable = 0) shows 0.8, that means that there is 80% likelihood that the correct outcome is imprisonment and 20% likelihood that the correct outcome is no imprisonment. In practice, since the social cost of wrongly imprisoning an accused is not symmetrical to the social cost of wrongly not imprisoning a guilty accused, we can set a higher threshold than 50% before sentencing an accused to imprisonment.

Linear regression, on the other hand, expresses its predicted values in numerical form. In linear regression, a prediction interval is 'an interval which uses the results of past sample to contain the results of a future sample from the same population with a specified probability'.³² A 95% prediction interval from a sample denotes that 95% of all future observations of the dependent variable will likely fall within this interval, given the same conditions in the input variables. These prediction intervals are given in the form of a pair of values representing the lower bound and upper bound of the prediction interval. On the other hand, it is not possible to give a probability statistic to the predicted value.

With this statistical background, it is difficult to understand what the reported 'recommended percentage' really refers to. One possibility is that the 'recommended percentage' is the predicted probability of logistic regression. If this were correct, then the predicted probability only tells us how likely the correct sentence will be, but tells us nothing about the accuracy of the recommended fines or terms of imprisonment.

5. Conclusion

The Magistrate's Court judgment in *PP v Denis P Modili* is laudable for clarifying certain aspects of the Artificial Intelligence in Court Sentencing (AiCOS) system used in Sabah and

³¹ Lim and Gong (n 3).

³² JK Patel, 'Prediction Intervals: A Review' (1989) 18 *Communications in Statistics—Theory and Methods* 2393 <<https://doi.org/10.1080/03610928908830043>>.

Sarawak. Due to the limited nature of the judgment, some technical information is not clear and not stated in the judgment. To date, the courts have not revealed the underlying model used to make predictions and there is no transparency to the whole set-up of the AI sentencing system. Nevertheless, it is hoped that with wider adoption of such systems in Malaysian courts, the relevant Bars will call for more public scrutiny of the accuracy and design of the AI sentencing systems in Malaysia in line with the global movement for explainable artificial intelligence.³³

Acknowledgement

The kind comments and assistance of Mohd. Munzil Muhamad, Siow-Hooi Tan and Roy W. L. Khong for clarifying some technical points are acknowledged. All errors remain the authors'.

Funding Information

The authors received no funding from any party for the research and publication of this article. ◆

³³ Pantelis Linardatos, Vasilis Papastefanopoulos and Sotiris Kotsiantis, 'Explainable AI: A Review of Machine Learning Interpretability Methods' (2020) 23 Entropy 18 <<https://dx.doi.org/10.3390/e23010018>>.