

Classification of Juvenile Delinquency Using Bayesian Network Learning: A Comparative Analysis

Abhinash Jenasamanta^{1,*} Subrajeet Mohapatra²

¹ Birla Institute of Technology, Mesra, India

² Birla Institute of Technology, Mesra, India

*Corresponding author. Email: samantajena99@outlook.com

ABSTRACT

The practice of engaging in offensive behavior on a frequent basis by a teenager is referred to as juvenile delinquency. Data mining and machine learning have been very effective techniques for a long time, allowing for efficient and accurate prediction in a variety of real-world applications. These techniques are gradually being implemented internationally in the area of criminal behavioral analysis, particularly in the detection of adolescent delinquency. According to studies, the risk of developing a deviant personality rises exponentially throughout the early period of adolescence. As a result, it makes perfect sense to identify deviant teenagers early and provide appropriate medical counselling. Providing routine psychological screening services for teenagers in a densely populated country is exceedingly difficult. Furthermore, due to a dearth of skilled clinicians, human evaluation of individual teenage behavior is highly subjective and time consuming. To handle this problem, an automated framework for the early identification of delinquent activity in juveniles has been implemented using Bayesian Network learning techniques. In this research, multi-class classification has been carried out using multiple Bayesian network based learning algorithms viz. K2 search, Simulated annealing, LAGD Hill Climbing and Tabu search. 5-Fold Cross-validation has been employed for multi class classification of juveniles into three groups based on severity levels viz. low, moderate and extreme. Simulation results are obtained and a comparative analysis shows that Bayesian Network with LAGD Hill Climber outperforms all other techniques.

Keywords: Juvenile delinquency, Bayesian network, Automated framework

1. INTRODUCTION

Data mining [1] is a technique for extracting meaningful information from massive data sets using automated techniques. It is fairly simple to build methods and procedures to categorize data for usage in real-world applications using statistics, artificial intelligence, and data warehousing. In contemporary research and for medical field applications, data mining has become a widely used technology. The most extensively utilized data mining techniques [2] for knowledge finding and pattern recognition are classification, regression and clustering. Machine learning (ML) [3] is the study of computer algorithms that improve themselves over time as a result of their utilization of data and experience. Artificial intelligence is considered to be an aspect of it. Machine learning algorithms create a model based on "training data" in order to make predictions or judgment without being specifically programmed to

do so. Bayesian Network [4] is being used in a variety of machine learning disciplines. It is a form of probabilistic analysis that has been integrated with various machine learning based applications.

Children under the age of eighteen [5] are considered juveniles. Juvenile delinquents are minors who commit crime consciously or innocently. Juvenile delinquency is a common psychological disorder among adolescents [6]. The behavior of juveniles gets disrupted by particular situations, which causes them to commit crimes. The juvenile justice system has devoted a great deal of time and effort to figure out what causes delinquency. The link between variables and outcomes is described by several frameworks. Researchers have determined that there is no unique cause for juvenile delinquency but combination of multiple risk factors often enhances a young person's chances of offending. Delinquency has become a global issue that affects

people from all walks of life. Delinquency is a means for adolescents to demonstrate false strength and bravery. Delinquency can also be used to exact retribution on other people or groups.

During the beginning of adolescence children develop the ability to reason. So there are chances of children entering into deviant activities at onset of adulthood. So, early detection of deviance behavior [7, 28] is of utmost importance. By using traditional psychological techniques [8] early detection and classification of adolescent delinquency is not feasible. In this research we propose automated classification of juvenile delinquency using different Bayesian Network search techniques. Delinquency risk assessment helps to determine the type of intervention that will best meet the needs of the youngster while also lowering his or her likelihood of offending. The doctor may recommend measures for the patient to lower his or her risk factors after this risk evaluation. Juvenile delinquency is caused by multiple factors [9], including individual variables, personal factors such as a lack of parental authority and monitoring, school-related factors such as a lack of school discipline and characteristics of friendship including deviant pals [10, 29]. Rathinabalan and Naaraayan(2017) conducted an annual study on the impact of different types of family influences on delinquency for adolescents staying in an observation house. To analyze the major risk factors that led to adolescent delinquency, they used multivariate logistic regression. Sixty juveniles and sixty school students were interviewed. Results revealed that father's behaviour as well as family structure were the major risk factors that led to delinquency among juveniles. Fernandez et al. (2016) analyzed the relation of school dropouts with different behavioral factors. Their results revealed that early dropouts indulged more in substance use than regular students. They discovered that inefficient parental control was a significant factor in the prediction of offending characteristics for dropouts. They recommended proper parental monitoring and effective school policies for preventing alcohol and substance use. Meldrum, Barnes, and Hay (2015) used information of neighbourhood, depression, parental factors, peer factors to develop regression models with latent factors which showed the connection between sleep and delinquency. The model proved that delinquency is indirectly related to lack of sleep where low self control plays the role of catalyst.

2. METHODOLOGY

For survey and analysis twelve schools of Ranchi District, India were chosen. As per past reports, the students are highly exposed to substance use and deviant peers. As a result, the likelihood of juvenile delinquency among students in these schools is much higher. The basic data came from interviews with adolescents in classes VII to XI at the designated schools. Deeper insights unto the individual's previous social activity in school was obtained by interviewing the corresponding teachers.

ISRD-3 (International Self-Report Delinquency Study questionnaire India Version) [11] was utilized for survey carried out in face to face format in initial stage. 138 students from age gap 11 to 17 participated in the survey. The head of each school (principal) was requested for permission to participate in the survey, followed by the students in the classroom. A group of three member team, including the authors and a behavioral psychologist, collected data in selected schools in the selected zone. The initial part of the process included an overview of the survey tool. Following that, the interview team met with the students for a face-to-face interview. The survey was completed by 138 students which included 76 boys and 62 girls with the responses being saved. The adolescents' response options for each individual aspect were converted into a numeric score using a variety of predetermined scales viz. 5-point Likert scales.

The complete questionnaire response set has been quantified into 47 psycho-behavioral features, according to the guidelines given by a dedicated panel of clinical and behavioral psychologists. Each feature's data field was transformed into a proportionate numeric entity. To measure various psycho-behavioral characteristics, the panel recommended that multiple conventional scale methods in psychology be utilized. Various other features, such as [12], group violence are also assessed, quantified, and numerically represented for quantitative analysis.

Before classification [13], feature scaling is needed to normalize/standardize various independent attributes into a fixed range. In this work Min-Max normalization [14] has been applied to normalize each of the 47 attributes in the data set within a range of zero and one. Low, moderate, and extreme risk levels are the three categories in which juveniles are classified depending on their levels of engaging in delinquent activity. In multi-class classification [15], using a separate model for each pair of classes

(called One vs One), instances have been categorized. The proposed multi-class classification architecture using multiple Bayesian Network based learning techniques, is used to classify juvenile behavioral risk levels into three groups for the early detection of deviance conduct in juveniles as depicted in Figure 1.

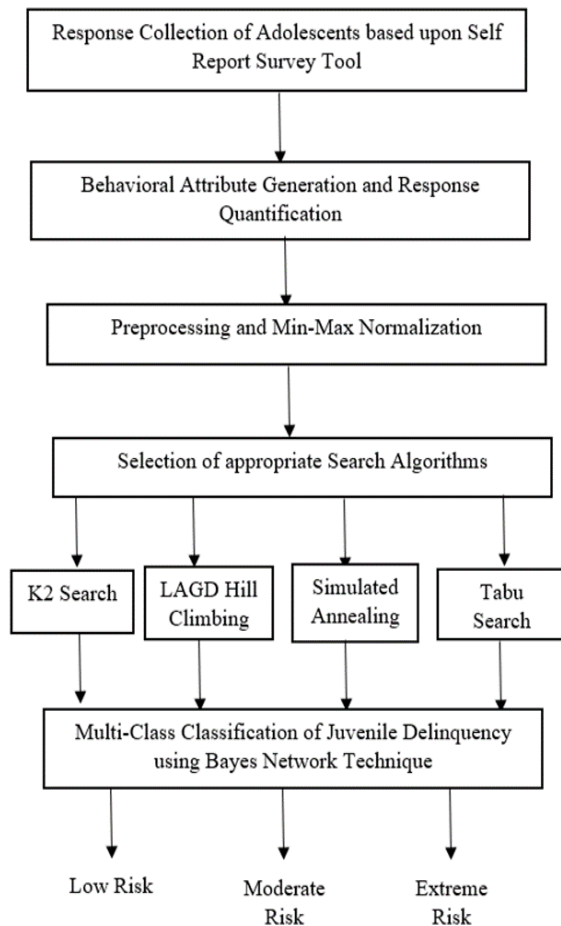


Figure 1 Proposed architecture for classification of delinquent behavior in juveniles using Bayesian Network Learning Technique.

3. BAYESIAN NETWORK LEARNING TECHNIQUES

Bayesian Network (BN) [16] are an effective tool for thinking in the face of uncertainty. BNs are able to handle domains with multiple variables because they provide a comprehensive and non-redundant representation of the underlying distribution. A Bayesian network is a probabilistic framework that uses a directed acyclic graph to depict a variable set and their conditional dependencies. Bayesian networks are perfect for forecasting the likelihood that any one of numerous possible known causes contributed to an event that occurred. A Bayesian

network indicates a directed acyclic graph where each node represents a unique variable and each edge represents a conditional dependency. The local Markov property asserts that given a node's parents, it is conditionally independent of its non-descendants. Bayesian networks satisfy this feature. A Bayesian network is a mathematical model [17] that depicts the joint probability distribution of a set of variables. The joint probability of $P(X,Y,Z)$ is defined as:

$$P(X,Y,Z) = P(X|Y) * P(Z|Y) * P(Y) \quad (1)$$

3.1 K2 Algorithm

The K2 algorithm [18] conducts a greedy search and does not restrict the count of parents for each node. This technique starts with the assumption of a node having zero parents and further proceeds for incremental addition of parent from a given sequence such that the increment maximizes the resulting structural score. Addition of parents on each node is halted when the count ceases to increment. This is expressed as a product of separate sub-scores, $g(X_i, P_i)$, for the group of variables and its parents, in the form:

$$P(G,D) = P(G) \prod_{i=1}^n g(X_i, P_i) \quad (2)$$

where $P(G)$ stands for the prior probability which is fixed for each G , for a random distribution D from the combined spread of X . This technique fetches the sequence which maximizes each feature.

3.2 LAGD Hill Climbing

LAGD Hill Climbing search technique [19] is a Look Ahead Hill Climbing algorithm implemented in Bayes Network. It evaluates a series of Look ahead sequence processes forming a model with highly improved efficiency as compared to the network obtained by implementing greedy techniques. It is used to compare the updated and old DAG (Directed acyclic Graphs). The look ahead steps (s) and processes in each look ahead steps (p) needs to be effectively calculated. The algorithm begins with initial DAG i with calculation of the gap (G) between new and old DAG's. The difference (G) between new and old DAG for all sequences are maximized.

3.3 Simulated Annealing

One of the most used metaheuristic methods for tackling optimization problems is the simulated annealing [20] algorithm. Simulated annealing may be applied to handle highly complex optimization

problems as well as many practical applications. The steps include establishing temperature minimization function $t(f)$ using required reduction techniques and mapping of temperature with time using annealing schedule. Finally the cost gap between the initial and final solution is calculated.

3.4 Tabu Search

Tabu Search (TS) [21] is a type of heuristic search technique that allows a heuristic search method to seek solution domain beyond local optimal range. The implementation of an adaptive memory, which allows for highly flexible search, is one of the primary advantages of this technique. Tabu Search operates on the concept of penalizing steps that accepts the result into past search spaces. Tabu search [22] takes into account non-improving outcomes in order to prevent local minima. The algorithm takes into account neighborhood solution, aspiration conditions and verification of termination conditions.

4. RESULTS

5-Fold Cross-Validation [23] has been carried out for evaluation of the Bayesian network. The model evaluation parameters for validation and performance analysis of various bayes network learning algorithms are Kappa statistics [24], AUC (ROC-Area) [25] and MCC value [26].

Accuracy = Correct prediction count / Total Predictions

True Positive Rate [27] is defined as:

$$TPR = TP / (TP + FN) \quad (3)$$

False Positive Rate is defined as:

$$FPR = FP / (FP + TN) \quad (4)$$

Where , FN and TN are False Negatives and True Negative respectively.

AUC is calculated as the area under ROC-Curve (TP Rate vs FP Rate). An Intel Core i5 CPU with 3.90 GHz system with 16 GB RAM configuration in Windows 10 environment has been utilized for simulation. The results of simulation carried out in the above technical specifications for the early detection of juvenile delinquency are presented. A comparative evaluation of classification performance of different bayes network based learning algorithms are described in Table 1 and Figure 2.

Table 1. Comparative Classification Performance of Multiple Bayes Network Based Learning Algorithms

Learning Algorithm	Classification Accuracy	Kappa Value	ROC-Area	MCC
K2 Search	78.02%	0.66	0.93	0.65
LAGD Hill Climbing	79.12%	0.68	0.93	0.67
Simulated Annealing	74.72%	0.62	0.89	0.61
Tabu Search	75.82 %	0.63	0.92	0.62

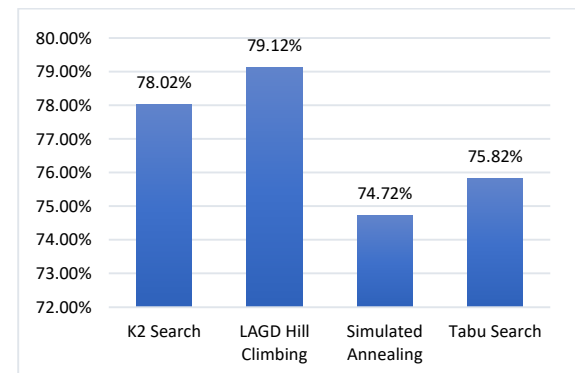


Figure 2 Comparison of classification accuracy of multiple Bayes Network learning algorithms for classification of juvenile delinquency

The results indicate that LAGD Hill Climbing search algorithm outperformed all other Bayes network learning algorithms for multi-class classification of juvenile delinquency with classification accuracy of 79.12% , Kappa value of 0.68 and ROC-Area of 0.93.

4.1 Comparative analysis of search algorithms using True Positive Rate (TPR) and False Positive Rate (FPR)

The comparative analysis of performance of multiple Naïve Bayes based learning algorithms in tabular as well as graphical format on basis of TPR and FPR is shown in Table 2:

Table 2. TPR and FPR for Classification of Juvenile Delinquency Using Bayesian Network

Learning Algorithm	TP Rate	FP Rate
K2 Search	0.78	0.12
LAGD Hill Climbing	0.79	0.12
Simulated Annealing	0.75	0.13
Tabu Search	0.76	0.14

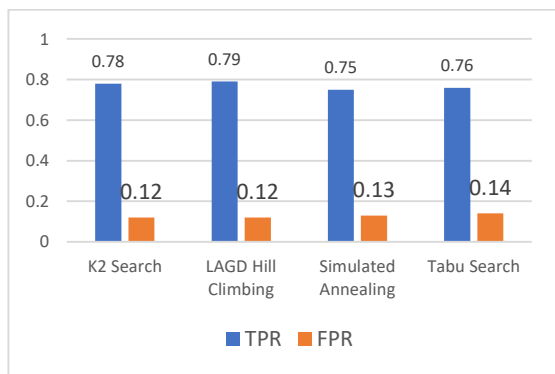


Figure 3 Comparison of classification performance of various Bayes Network learning algorithms using TPR and FPR

The achieved TPR and FPR values indicate that LAGD Hill climbing search algorithm outperforms all other techniques in classifying juvenile delinquency using Bayesian Network.

5. CONCLUSION

Early diagnosis of deviant behavior in juveniles is a major challenge for the society. In this study, different Bayes network based search approaches were implemented for multi-class classification of juvenile delinquency and LAGD Hill climbing search algorithm emerged as the best search technique. In future, prediction/classification models used for assessing juvenile delinquency needs to adapt to current trends in technology in order to provide more effective and accurate prediction as well as assist in development of robust automated diagnosis framework. We are hopeful that our analysis of juvenile delinquency may provide detailed insights to researchers conducting advanced quantitative research in this domain. In future, advanced machine learning algorithms viz. SVM kernel learning, Multi-Layer Perceptron (MLP), ensemble learning and deep learning may be utilized to handle high dimensional behavioral and clinical data to identify juvenile delinquency and its causes with better accuracy and performance. The sequencing of data are crucial to many prediction models such as classification and clustering. As a result, effective data repository management saves time and money while also assisting in the successful aggregation of data from disparate data sources.

AUTHORS' CONTRIBUTIONS

The study was planned and designed by all authors. Abhinash Jenasamanta and Subrajeet Mohapatra contributed to experimental design, data collection, data analysis and simulation. All authors

conducted the article revision and approved the final manuscript.

ACKNOWLEDGMENTS

The authors are thankful to Professor Ineke Haen Marshall, School of Criminology and Criminal Justice, Northeastern University, Boston for providing us with the required permissions for analysing and using ISRD-3 questionnaire. The authors are thankful to Cognitive Science Research Initiative (CSRI), Department of Science & Technology (DST), Government of India for funding this research.

REFERENCES

- [1] Koedinger K.R., D'Mello S., McLaughlin E.A., Pardos, Z.A., and Rose C.P.: Data mining and education. Wiley Interdisciplinary Reviews: Cognitive Science, vol. 6, pp. 333-353, (2015). DOI: 10.1002/wcs.1350
- [2] Tan P.N., Steinbach M., and Kumar V.: Introduction to data mining. Pearson Education India, (2016).
- [3] Zhang X.D.: A matrix algebra approach to artificial intelligence. Springer, (2020).
- [4] Jiang W., Cao Y. and Deng X.: A novel Z-network model based on Bayesian network and Z-number. IEEE Transactions on Fuzzy Systems, vol. 28, pp. 1585-1599, (2019). DOI: 10.1109/TFUZZ.2019.2918999
- [5] Lundman R.J.: Prevention and control of juvenile delinquency. Oxford University Press, (2001).
- [6] Shoemaker D.J.: Juvenile delinquency. Rowman & Littlefield, (2017).
- [7] Bobbio A., Arbach K. and Illescas S.R.: Juvenile delinquency risk factors: Individual, social, opportunity or all of these together?. International Journal of Law, Crime and Justice, vol. 62, p. 100388, (2020). DOI: <https://doi.org/10.1016/j.ijl cj.2020.100388>
- [8] Kratcoski P.C., Kratcoski L.D. and Kratcoski P.C.: Juvenile delinquency: Theory, research, and the juvenile justice process. Springer, (2020).

- [9] Nisar M., Ullah S., Ali M. and Alam S.: Juvenile delinquency: The Influence of family, peer and economic factors on juvenile delinquents. *Applied Science Reports*, vol. 9, pp. 37-48, (2015).
- [10] Marshall I.H., Wills C. and Marshall C.E.: Parents Who Hit, Troubled Families, and Children's Happiness: Do Gender and National Context Make a Difference?. In *Crime Prevention and Justice in 2030*, Springer, Cham, pp. 199-230, (2021). DOI: https://doi.org/10.1007/978-3-030-56227-4_10
- [11] Waugh M.H., McClain C.M., Mariotti E.C., Mulay A.L., DeVore E.N., Lenger, K.A. Russell, A.N. Florimbio, A.R. Lewis, K.C. Ridenour, J.M. and Beevers, L.G.: Comparative content analysis of self-report scales for level of personality functioning. *Journal of personality assessment*, vol. 103, pp. 161-173, (2021).
- [12] Hughes L.A., Schaible L.M. and Kephart T.: Gang graffiti, group process, and gang violence. *Journal of quantitative criminology*, pp.1-20, (2021).
- [13] Yang C., Sun Y., Ladubec C. and Liu Y.: Developing machine learning-based models for railway inspection. *Applied Sciences*, vol. 11, p. 13, (2021).
- [14] Kim H.J., Baek J.W. and Chung K.: Associative Knowledge Graph Using Fuzzy Clustering and Min-Max Normalization in Video Contents. *IEEE Access*, vol. 9, pp. 74802-74816. (2021). DOI: 10.1109/ACCESS.2021.3080180
- [15] Mortier T., Wydmuch M., Dembczyński K., Hüllermeier E. and Waegeman W.: Efficient set-valued prediction in multi-class classification. *Data Mining and Knowledge Discovery*, pp. 1-35, (2021).
- [16] Marcot B.G. and Penman T.D.: Advances in Bayesian network modelling: Integration of modelling technologies. *Environmental modelling & software*, vol. 111, pp. 386-393, (2019).
- [17] Sun L. and Erath A.: A Bayesian network approach for population synthesis. *Transportation Research Part C: Emerging Technologies*, vol. 61, pp. 49-62, (2015).
- [18] Lerner B. and Malka* R.: Investigation of the K2 algorithm in learning Bayesian network classifiers. *Applied Artificial Intelligence*, vol. 25, pp. 74-96, (2011). DOI: <https://doi.org/10.1080/08839514.2011.529265>
- [19] Khasanah A.U.: A comparative study to predict student's performance using educational data mining techniques. In *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 215, p. 012036, (2017).
- [20] Lee S. and Kim S.B.: Parallel simulated annealing with a greedy algorithm for Bayesian network structure learning. *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, pp. 1157-1166, (2019). DOI: 10.1109/TKDE.2019.2899096
- [21] Liu H., Zhang J., Zhang X., Kurniawan A., Juhana T. and Ai B.: Tabu-search-based pilot assignment for cell-free massive MIMO systems. *IEEE Transactions on Vehicular Technology*, vol. 69, pp. 2286-2290, (2019).
- [22] Glover F.: Tabu search—part I. *ORSA Journal on computing*, vol. 1, pp.190-206, (1989).
- [23] Marcot B.G. and Hanea A.M.: What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?. *Computational Statistics*, vol. 36, pp. 2009-2031, (2021). DOI: <https://doi.org/10.1007/s00180-020-00999-9>
- [24] Chalak L.F., Pavageau L., Huet B. and Hynan L.: Statistical rigor and kappa considerations: which, when and clinical context matters. *Pediatric research*, vol. 88(1), pp. 5-5, (2020).
- [25] Bowers A.J. and Zhou X.: Receiver operating characteristic (ROC) area under the curve (AUC): a diagnostic measure for evaluating the accuracy of predictors of education outcomes. *Journal of Education for Students Placed at Risk (JESPAR)*, vol. 24, pp. 20-46, (2019). DOI: <https://doi.org/10.1080/10824669.2018.1523734>
- [26] Chicco D. and Jurman G.: The advantages of the Matthews correlation coefficient (MCC)

- over F1 score and accuracy in binary classification evaluation. *BMC genomics*, vol. 21, pp. 1-13, (2020).
- [27] Bae M., Kim H.I., Park J.H., Ryu B.H., Chang J., Sung H., Jung J., Kim M.J., Kim S.H., Lee S.O. and Choi S.H.: Improvement of blood culture contamination rate, blood volume, and true positive rate after introducing a dedicated phlebotomy team. *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 38, pp. 325-330, (2019). DOI: <https://doi.org/10.1007/s10096-018-3430-4>
- [28] D Suar, J Gochhayat, Influence of biological sex and gender roles on ethicality, *Journal of Business Ethics* 134 (2), (2016) 199-208.
- [29] S Sarkar, S Samanta, A Mitra, Economic Victimization of convict's family: Statistical analysis through SPSS, *International Journal of Knowledge-based and Intelligent Engineering Systems*, 24(3) (2020) 261-268.