Predicting the offender
Frequency versus bayes

Sutmuller, August Daniel; Den Hengst, Marielle; Barros, Ana Isabel; Van der Vecht, Bob; Noordkamp, Wouter; Van Gelder, P. H.A.J.M.

DOI
10.1109/EISIC49498.2019.9108891

Publication date
2019

Document Version
Final published version

Published in
Proceedings of the 2019 European Intelligence and Security Informatics Conference, EISIC 2019

Accepted Author Manuscript. Link to published article (IEEE): https://doi.org/10.1109/EISIC49498.2019.9108891

Predicting the Offender: Frequency versus Bayes

August Daniel Sutmuller
Faculty of Technology, Policy and Management
Delft University of Technology
Delft, The Netherlands
a.d.sutmuller@tudelft.nl

Mariëlle den Hengst
RTI Laboratories
Dutch Police
The Hague, The Netherlands

Ana Isabel Barros
TNO
Defence, Safety and Security
The Hague, The Netherlands

Bob van der Vecht
TNO
Defence, Safety and Security
The Hague, The Netherlands

Wouter Noordkamp
TNO
Defence, Safety and Security
The Hague, The Netherlands

P.H.A.J.M. van Gelder
Faculty of Technology, Policy and Management
Delft University of Technology
Delft, The Netherlands

***Abstract*— In this paper two Bayesian approaches and a frequency approach are compared on predicting offender output variables based on the input of crime scene and victim variables. The K2 algorithm, Naïve Bayes and frequency approach were trained to make the correct prediction using a database of 233 solved Dutch single offender/single victim homicide cases and validated using a database of 35 solved Dutch single offender/single victim homicide cases. The comparison between the approaches was made using the measures of overall prediction accuracy and confidence level analysis. Besides the comparison of the three approaches, the correct predicted nodes per output variable and the correct predicted nodes per validation case were analyzed to investigate whether the approaches could be used as a decision tool in practice to limit the incorporation of persons of interest into homicide investigations. The results of this study can be summarized as: the non-intelligent frequency approach shows similar or better results than the intelligent Bayesian approaches and the usability of the approaches as a decision tool to limit the incorporation of persons of interest into homicide investigations should be questioned.**

Keywords— offender, profiling, homicide, frequency, Bayesian

I. Introduction

Nowadays our personal data is stored and easily accessible in databases ranging from the registration of the city where we live, the sports club we join, the school we attend and the shops we visit. And, not only our membership leads to the storage of our data, in this era we are recorded by CCTV and carry goods that leave traces, like our smartphone that places us on certain places at certain times and our car that is registered entering and leaving cities or highways. These developments enable the police to incorporate lots of persons of interest around a crime scene. Sutmuller, den Hengst, Barros & van Gelder [12] showed that the incorporation of the actual perpetrator into a homicide investigation can be accelerated if categories of persons of interest are incorporated. Categories that were used: people with a relationship with the victim, people that were present in a certain geographical area, people with certain physical characteristics, people with previous convictions and people owing certain registered goods. The downside of using categories is the incorporation of a great number of persons of interest [12]. The increased number of persons of interest makes the identification of the actual perpetrator a daunting task. It would be very helpful when crime scene and victim characteristics, known on the first day of a criminal investigation, can be used to steer the incorporation of persons of interest in that particular investigation. Since the 1970s profilers tend to use crime scene and victim characteristics to build profiles with characteristics of the perpetrator. These profiles can be used as an investigative tool to prioritize suspects [10].

Criminal profiling has its origin in the 1970s with the establishment of Behavioral Science Unit (BSU) of the Federal Bureau of Investigation (FBI) in Quantico. Employees of the BSU started interviewing serial killers, trying to get inside their head and understand them. This resulted in the first articles in 1980 [6]. In one of the articles, the well-known distinction between the organized and disorganized offender was introduced [9]. Characteristics of the crime scene point in the direction of an offender type. The profile of this offender type consists of characteristics of the offender. The inference from crime scene actions to offender characteristics is the core of criminal profiling [3]. And, although no empirical evidence was found for the organized and disorganized distinction [4] and offender profiling in general is still under debate, a growing interest for criminal profiling is seen over the past decades [7]. The main criticism on qualitative offender profiling, based on behavioral science, is focused on two issues; the lack of empirical evidence that congruency exists between the way a murderer commits a crime and who he is and the lack of taking the influence of the situational aspects into account[8].

With the upswing of data processing possibilities and the expansion of stored data to use, more statistical approaches of criminal profiling started to appear. The use of statistical datasets to aid the investigation of major crimes is referred to as quantitative profiling. A study of Baumgartner, Ferrari and Salfati [2] can be seen as a good example of quantitative profiling. They used a Bayesian network (BN) modeling approach on a database of solved cases, the aim was to seek non-obvious and valuable patterns between variables resulting in a model that could be used as a decision tool. From a training database of 200 solved single offender/single victim cases, a model was obtained that used 36 crime scene input variables and 21 offender output variables. These variables were selected by investigators, criminologists, and forensic psychologists. The model was validated using 47 cases in which the 21 offender variables were predicted by the 36 crime scene variables. In the study two algorithms and the frequency approach were compared. The results support the idea that underlying patterns exist between offenders and their crime but the usefulness of the algorithms over frequency was not apparent [2].

In the era where all data is digitally available, it would be interesting to investigate whether data from historical crime scenes and their accessory perpetrator could help us to steer the incorporation of persons of interest into homicide investigations. In this study we analyze a database of 268 recently solved Dutch single offender/single victim cases. The aim is to compare two Bayesian approaches (K2 algorithm and Naïve Bayes) with the frequency approach on predicting offender output variables based on the input of crime scene and victim variables and to provide a decision tool that could be used in practice to limit the incorporation of persons of interest into homicide investigations.

II. PROCEDURE

A. Database

A database of 268 solved Dutch single offender/single victim cases committed between 2011 and 2018 was constructed. The included cases are homicides in this period meeting the single offender/single victim criterion and were marked as solved by the Dutch police. The marking means that a suspect is handed over to the public prosecutor to be brought to trial. In 238 (89%) cases the offender and 135 cases (50%) the victim was male. The causes of death were gunshot wounds in 32 cases and stabbing injuries in 138 cases. In 78 cases manual injuries (hitting, kicking and strangulation) were found on the victim's body and in 31 cases a blunt instrument was used. In 8 cases the homicide was categorized as a sexual crime, in 35 cases the perpetrator had a criminal motive and in 210 cases the victim and perpetrator shared the same social network. The cases were solved after 1 to 572 days with a mean of 16.14 days. The database was divided into a training set and a validation set. For validation, the 35 most recently solved cases were selected because the aim is to use historical cases to steer new and ongoing investigations. The other 233 solved Dutch single offender/single victim cases were used to train the three approaches for making the correct prediction.

B. Variables

The 36 crime scene input variables and 21 offender output variables, as used by Baumgartner, Ferrari and Salfati [2], were used and extended with fifteen victim characteristics about the criminal history, age, social status and sex of the victim and four offender characteristics about the relationship between the victim and the offender. All variables are discrete. With the added victim and offender characteristics, 51 crime scene (XI) and victim (XV) input variables[1] and 25 offender (XO) output variables[2] were included in the analysis. The 76 variables were scored either Yes ("Y") or No ("N"), when no information was available the variable was scored as Unknown ("U").

C. Approaches

In this study three different approaches are used to handle the data: the K2 algorithm, Naïve Bayes and frequency. The existence of missing data on a number of variables in our database and the fact that the K2 algorithm requires that no missing data exists, made that we had to adopt approaches to deal with this problem. This resulted in two different models. Model 1: Consider only the input and output variables that do not have the value "U" in any of the cases. This leads to a model with 20 input variables and 14 output variables. Model 2: Take into account all input and output variables that have at most 10% unknowns. All unknowns are treated as "N". This leads to a model with 45 input variables and 19 output variables. In order to be able to compare the three approaches, Model 1 and 2, as used by the K2 algorithm, are used for the Naïve Bayes and frequency approach.

---

[1] XI1: Foreign object penetration XI2: Face deliberately hidden XI3: Victim was blindfolded XI4: Wounds caused by blunt instrument XI5: Suffocation (other than strangulation) XI6: Vaginal penetration XI7: Anal penetration XI8: Face up (victim found as they fell) XI9: Victim partially undressed XI10: Victim naked XI11: Deliberate clothing damaged XI12: Bound (at one point) XI13: Stabbing injuries XI14: Manual injuries (hitting, kicking, strangled) XI15: Gunshot wounds XI16: Wounds to the head XI17: Wounds to the face XI18: Wounds to the neck XI19: Wounds to the torso XI20: Wounds to the limbs XI21: Multiple wounds to one body area (MWOA) XI22: Multiple wounds distributed across different body parts XI23: Weapon brought to scene XI24: Weapon from the scene XI25: Identifiable property stolen (identification property) XI26: Non-identifiable property stolen (non-valuable and unidentifiable) XI27: Valuable property stolen XI28: Body hidden (outside) XI29: Body transported XI30: Offender forensically aware XI31: Victim found at the same scene where they were killed XI32: Sexual crime XI33: Arson to crime scene/body XI34: Victim found in water XI35: Victim drugged and/or poisoned XI36: Victim covered (i.e., inside rather than outside) XV1: Young victim under 21 years XV2: Criminal record of theft XV3: Criminal record of fraud XV4: Criminal record of burglary XV5: Unemployed at the time of offense XV6: Male XV7: Criminal record of violence XV8: Criminal record of committing damage XV9: Criminal record of disorderly conduct XV10: Record of imprisonment XV11: Sexual related criminal record XV12: Armed services, past or present XV13: History of abusiveness in past relationships XV14: Attempts of suicide XV15: Psychiatric disorders

[2] XO1: Young offender between 17-21 years XO2: Criminal record of theft XO3: Criminal record of fraud XO4: Criminal record of burglary XO5: Relationship with victim XO6: Unemployed at the time of offense XO7: Male XO8: Familiar with area of offense occurrence XO9: Criminal record of violence XO10: Criminal record of committing damage XO11: Criminal record of disorderly conduct XO12: Record of imprisonment XO13: Sexual related criminal record XO14: Armed services, past or present XO15: Knew victim XO16: History of abusiveness in past relationships XO17: Attempts of suicide XO18: Psychiatric disorders XO19: Related to victim XO20: Blood relative to victim XO21: Turned self into police X022: Criminal relation between offender and victim XO23: Geographical relation between offender and crime scene XO24: Offender was part of social network of the victim XO25: Other relationship between victim and offender

The K2 algorithm is used to construct a Bayesian model, i.e. the directed acyclic graph that defines the relation between the variables. Given this graph, the conditional probabilities of the network can be estimated based on the 233 cases. For the validation of the model, for each case in the set of 35 validation cases, the input variables are entered as evidence in the network. Next, the probability for the value of each of the output variables is determined. If the probability of "Y" is 50% or more, the model predicts "Y", otherwise the model predicts "N". The resulting prediction of the output variables is compared to the value of the output variables of the validation case under consideration. The BayesNet Toolbox for Matlab is used to construct the model, to calculate the conditional probabilities (function learn_params) and to predict the output variables (function pearl_inf_engine). Because the ordering of input variables could influence the predictions by the K2 algorithm, four orderings were considered. The results of the orderings were comparable, a mean score of the four orderings is given in the results section. Fig. 1 shows an example of a network created by K2.

In the Naïve Bayes approach every output variable is estimated individually based on all input variables. As Model 1 uses 20 input variables and 14 output variables, this results in 14 networks where 20 inputs are linked to the output variable. In Model 2 a total of 45 inputs are linked to the 19 output variables. All links in the network have a weight parameter defining the strength of the input variable to predict the output value. This weight is known as the Bayes factor (BF), and is directly calculated from the conditional probabilities from the training dataset. The presence of an input factor has a different strength than the absence of the factor. Fig. 2 shows the network structure using Naïve Bayes.

Frequency predicts the output values without looking at any input variable. When in more than 50% of the cases in the training set, the output value occurred, this output value is predicted for all cases in the validation database.

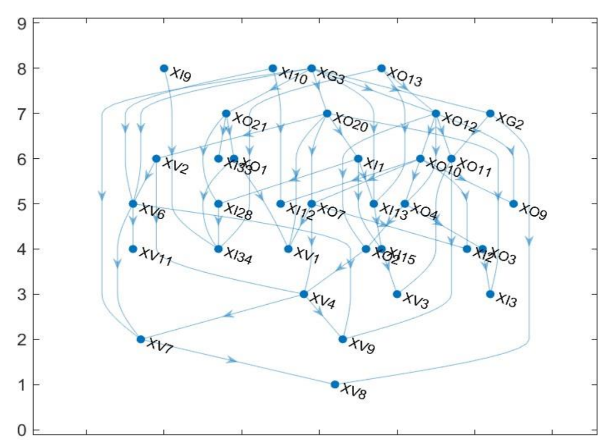

Fig. 1. Example of network created by the K2 algorithm.

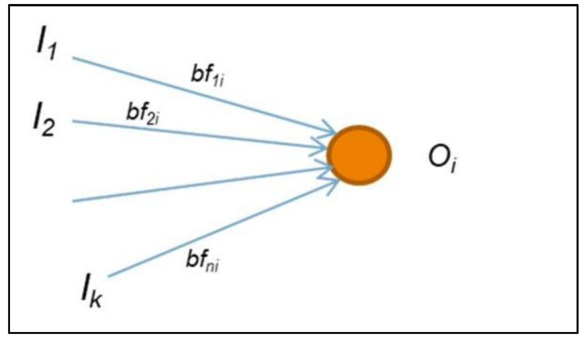

Fig. 2. Network structure using Naïve Bayes.

D. *Measures*

In order to make a comparison between the three approaches, two measures are used: overall prediction accuracy (OPA) and confidence level analysis (CLA). OPA is defined as the frequency at which output variables are inferred correctly over the 35 validation cases. A predicted variable is said to be inferred correctly, or its prediction is said to be correct, when the true (observed) state is equal to the predicted value. The OPA is the percentage of correct predictions over the total number of predictions [2]. Further comparison of the K2 algorithm and the Naïve Bayes model involves the confidence level of each prediction. The accuracy of nodes predicted with a confidence level (CL) is denoted by CLA and is calculated by the following formula:

$$CLA = KC,CL/KCL * 100$$

where, KC,CL is the total number of correct predictions with a specified confidence level, and KCL is the total number of nodes in the specified confidence level. To compare the CLA of the Bayesian approaches to the non-intelligent frequency approach, the frequency of occurrence of a variable in the training set is used. For example, 93 out of 200 training cases involve an offender with a prior theft conviction, which leads to 46.5% offenders with a prior theft conviction. These probabilities acquired from the training database can be interpreted as the confidence levels of the frequency approach [2].

In order to gain more insight in the performance of the approaches, the correct predicted nodes per output variable and the correct predicted nodes per validation case are analyzed. The analysis of individual nodes is conducted to investigate the usability of the approaches as a decision tool.

III. Results

In this section, the results of the three approaches are presented. The OPA shows the percentage of correct predictions over the total number of predictions. A total number of 490 predictions were made in Model 1 and 665 predictions were made in Model 2. The total of correct predictions by the three approaches is given in Table I and Table II. In both models the frequency approach outperforms the Bayesian approaches.

Table III and Table IV show the results of the CLA. Frequency and the K2 algorithm show similar results in the distribution of nodes over confidence levels and the percentage of correct predicted nodes in these levels for both models. The Naïve Bayes approach shows less predictions with high confidence levels and the percentage of correct predicted nodes in the different confidence levels is, except for >80% and >90% in Model 1, lower than the percentages of the other two approaches.

TABLE I. Overall prediction accuracy of the three approaches for the total of 490 predictions in Model 1.

	K2 algorithm	**Naïve Bayes**	**Frequency**
OPA (%)	74.08%	73.27%	75.10%
Correct predictions	363	359	368

TABLE II. Overall prediction accuracy of the three approaches for the total of 665 predictions in Model 2.

	K2 algorithm	**Naïve Bayes**	**Frequency**
OPA (%)	72.18%	66.17%	76.69%
Correct predictions	480	440	510

TABLE III. The total number of nodes in the different confidence levels (Kcl), the correct predictions within these confidence levels (Kc, cl) and accuracy within the confidence levels (CLA) are displayed for the three approaches in Model 1.

	The K2 algorithm | Naïve Bayes | Frequency		
CL (%)	***Kcl***	***Kc,cl***	***CLA (%)***
≥50%	420 | 490 | 455	331 | 359 | 355	78.8 | 73.3 | 78.0
≥60%	350 | 455 | 385	292 | 339 | 318	83.4 | 74.5 | 82.6
≥70%	350 | 280| 350	292 | 223 | 297	83.4 | 79.6 | 84.9
≥80%	245 | 105 | 245	212 | 92 | 217	86.5 | 87.6 | 88.6
≥90%	105 | 35 | 105	97 | 32 | 98	92.4 | 91.4 | 93.3

TABLE IV. The total number of nodes in the different confidence levels (Kcl), the correct predictions within these confidence levels (Kc, cl) and accuracy within the confidence levels (CLA) are displayed for the three approaches in Model 2.

	The K2 algorithm | Naïve Bayes | Frequency		
CL (%)	***Kcl***	***Kc,cl***	***CLA (%)***
≥50%	595 | 630 | 630	450 | 425 | 497	75.6 | 67.5 | 78.9
≥60%	455 | 490 | 560	374 | 348 | 460	82.2 | 71.0 | 82.1
≥70%	420 | 210 | 490	353 | 166 | 415	84.0 | 79.0 | 84.7
≥80%	280 | 140 | 350	247 | 114 | 308	88.2 | 81.4 | 88.0
≥90%	105 | 0 | 140	97 | 0 | 130	92.4 | 0.0 | 92.9

Table V and Table VI show the predictive value of the individual output variables. Two output variables (male and social network victim) in Model 1 and three output variables (male, knew victim and social network victim) in Model 2 lead to more “Y” predictions for all three approaches. The Naïve Bayes approach shows four more output variables with a “Y” prediction in Model 2. The distribution between true positives and true negatives contributing to the OPA shows that the true negatives represent a large share of the correct predictions for all three approaches.

Table VII and Table VIII show the percentage of correct predictions per validation case and the percentage of these correct predictions that was contributed by true negatives. When the approaches are compared one can state that the K2 (Model 1: M = 74%, SD 14%, Min. = 36%, Max. = 100 %; Model 2: M = 72%, SD 13%, Min. = 37%, Max. = 100 %) and frequency (Model 1: M = 75%, SD 16%, Min. = 36%, Max. = 100 %; Model 2: M = 77%, SD 12%, Min. = 42%, Max. = 100 %) show similar results and Naïve Bayes (Model 1: M = 73%, SD 17%, Min. = 36%, Max. = 100 %; Model 2: M = 66%, SD 13%, Min. = 42%, Max. = 95%) seems to scores less, although not significantly. The percentage of correct predicted nodes per validation case heavily depend on true negatives for all three approaches. The Naïve Bayes (Model 1: 72%, Model 2: 68%) approach shows the smallest share of true negatives. The K2 algorithm approach (Model 1: 82%, Model 2: 83%) and frequency (Model 1: 84%, Model 2: 84%) show great dependency of true negatives in the correct predictions per validation case.

TABLE V. THE PREDICTED VALUE, TRUE POSITIVES (TP), FALSE POSITIVES (FP), TRUE NEGATIVES (TN), FALSE NEGATIVES (FN) AND PERCENTAGE OF CORRECT PREDICTIONS AS PREDICTED BY THE THREE APPROACHES ARE DISPLAYED FOR ALL INDIVIDUAL OUTPUT VARIABLES FOR MODEL 1. CR STANDS FOR CRIMINAL RECORD.

	The K2 algorithm | Naïve Bayes | Frequency						
Output variable	***Yes***	***No***	***TP***	***FP***	***TN***	***FN***	***Correct (%)***
offender 17-21	2 | 3 | 0	33 | 32 | 35	0 | 1 | 0	2 |2 |0	29 |29 |31	4 |3 |4	82.9 |85.7 |88.6
cr of theft	4 | 14 | 0	31 | 21 | 35	1 |10 | 0	3 |4 |0	15 |14 |18	16 |7 |17	45.7 |68.6 |51.4
cr of fraud	0 | 8 | 0	35 | 27 | 35	0 |1 |0	0 |7 |0	33 |26 |33	2 |1 |2	94.3 |77.1 |94.3
cr of burglary	0 | 7 | 0	35 | 28 | 35	0 |2 |0	0 |5 |0	27 |22 |27	8 |6 |8	77.1 |68.6 |77.1
male	**34 | 32 | 35**	**1 | 3 | 0**	**30 |29 |31**	**4 |3 |4**	**0 |1 |0**	**1 |2 |0**	**85.7 |85.7 |88.6**
cr of violence	3 | 17 | 0	32 | 18 | 35	2 |9 |0	1 |8 |0	18 |11 |19	14 |7 |16	57.1 |57.1 |54.3
cr of committing damage	0 | 9 | 0	35 | 26 | 35	0 |2 |0	0 |7 |0	28 |21 |28	7 |5 |7	80.0 |65.7 |80.0
cr disorderly conduct	6 | 15 | 0	29 | 20 | 35	3 |10 |0	3 |5 |0	18 |16 |21	11 |4 |14	60.0 |74.3 |60.0
record of imprisonment	5 | 14 | 0	30 | 21 | 35	3 |11 |0	2 |3 |0	11 |10 |13	19 |11 |22	40.0 |60.0 |37.4
sexual related cr	1| 5| 0	34 | 30 | 35	0 |1 |0	1 |4 |0	28 |25 |29	6 |5 |6	80.0 |74.3 |82.9
blood relative to victim	1 | 1 | 0	34 | 34 | 35	0 |0 |0	1 |1 |0	32 |32 |33	2 |2 |2	91.4 |91.4 |94.3
turned into police	0 | 2 | 0	35 | 33 | 35	0 |1 |0	0 |1 |0	27 |26 |27	8 |7 |8	77.1 |77.1 |77.1
geography	0 | 8 | 0	35 | 27 | 35	0 |0 |0	0 |8 |0	32 |24 |32	3 |3 |3	91.4 |68.6 |91.4
social network victim	**35 | 28 | 35**	**0 | 7 | 0**	**26 |22 |26**	**9 |6 |9**	**0 |3 |0**	**0 |4 |0**	**74.3 |71.4 |74.3**
Total	91 | 163 | 70	399 | 327 | 420	65 |99 |57	26 |64 |13	298 |260 |311	101 |67 |109	74.1 |73.3 |75.1

TABLE VI. THE PREDICTED VALUE, TRUE POSITIVES (TP), FALSE POSITIVES (FP), TRUE NEGATIVES (TN), FALSE NEGATIVES (FN) AND PERCENTAGE OF CORRECT PREDICTIONS AS PREDICTED BY THE THREE APPROACHES ARE DISPLAYED FOR ALL INDIVIDUAL OUTPUT VARIABLES FOR MODEL 2. CR STANDS FOR CRIMINAL RECORD.

	The K2 algorithm | Naïve Bayes | Frequency						
Output variable	***Yes***	***No***	***TP***	***FP***	***TN***	***FN***	***Correct (%)***
offender 17-21	1 |9 |0	34 |26 |35	0 |3 |0	1 |6 |0	30 |25 |31	4 |1 |4	85.7 |80.0 |88.6
cr of theft	4 |15 |0	31 |20 |35	1 |10 |0	3 |5 |0	15 |13 |18	16 |7 |17	45.7 |65.7 |51.4
cr of fraud	0 |17 |0	35 |18 |35	0 |2 |0	0 |15 |0	33 |18 |33	2 |0 |2	94.3 |57.1 |94.3
cr of burglary	1 |11 |0	34 |24 |35	0 |4 |0	1 |7 |0	26 |20 |27	8 |4 |8	74.3 |68.6 |77.1
relationship with victim	19 |10 |0	16 |25 |35	7 |6 |0	12 |4 |0	12 |20 |24	4 |5 |11	54.3 |74.3 |68.6
male	**29 |25 |35**	**6 |10 |0**	**26 |22 |31**	**3 |3 |4**	**1 |1 |0**	**5 |9 |0**	**77.1 |65.7 |88.6**
cr of violence	5 |27 |0	30 |8 |35	2 |13 |0	3 |14 |0	16 |5 |19	14 |3 |16	51.4 |54.4 |54.3
cr of committing damage	1 |13 |0	34 |22 |35	1 |4 |0	0 |9 |0	28 |19 |28	6 |3 |7	82.9 |65.7 |80.0
cr of disorderly conduct	5 |23 |0	30 |12 |35	2 |12 |0	3 |11 |0	18 |10 |21	12 |2 |14	57.1 |62.9 |60.0
record of imprisonment	5 |20 |0	30 |15 |35	3 |15 |0	2 |5 |0	11 |8 |13	19 |7 |22	40.0 |65.7 |37.1
sexual related cr	0 |14 |0	35 |21 |35	0 |2 |0	0 |12 |0	29 |17 |29	6 |4 |6	82.9 |54.3 |82.9
knew victim	**27 |18 |35**	**8 |17 |0**	**20 |16 |27**	**5 |2 |6**	**1 |4 |0**	**7 |11 |0**	**60.0 |57.1 |77.1**
related to victim	1 |8 |0	34 |27 |35	0 |2 |0	1 |6 |0	31 |26 |32	3 |1 |3	88.6 |80.0 |91.4
blood relative to victim	1 |7 |0	34 |28 |35	0 |1 |0	1 |6 |0	32 |27 |33	2 |1 |2	91.4 |80.0 |94.3
turned into police	0 |9 |0	35 |26 |35	0 |2 |0	0 |7 |0	27 |20 |27	8 |6 |8	77.1 |62.9 |77.1
criminal	4 |10 |0	31 |25 |35	1 |6 |0	3 |4 |0	25 |24 |28	6 |1 |7	74.3 |85.7 |80.0
geography	0 |19 |0	35 |16 |35	0 |1 |0	0 |18 |0	32 |14 |32	3 |2 |3	91.4 |42.9 |91.4
social network victim	**26 |21 |35**	**9 |14 |0**	**18 |19 |26**	**8 |2 |9**	**1 |7 |0**	**8 |7 |0**	**54.3 |74.3 |74.3**
other	0 |11 |0	35 |24 |35	0 |1 |0	0 |10 |0	31 |21 |31	4 |3 |4	88.6 |62.9 |88.6
Total	129 |287 |105	536 |378 |560	81 |141 |84	46 |146 |19	399 |299 |426	137 |774 |134	72.2 |66.2 |46.7

TABLE VII. THE PERCENTAGE OF CORRECT PREDICTED NODES PER VALIDATION CASE AND THE PERCENTAGE OF TRUE NEGATIVES CONTRIBUTING TO THE CORRECT PREDICTED NODES IN MODEL 1.

	The K2 algorithm	**Naïve Bayes**	**Frequency**
Case	***Correct per case (%) | True negatives (%)***		
1	93 | 85	86 | 83	93 | 85
2	79 | 82	79 | 82	79 | 82
3	64 | 78	79 | 55	57 | 88
4	86 | 83	50 | 71	86 | 83
5	71 | 80	71 | 80	71 | 80
6	79 | 82	79 | 82	79 | 82
7	70 | 79	36 | 20	71 | 80
8	100 | 86	100 | 86	100 | 86
9	86 | 83	86 | 83	86 | 83
10	64 | 78	86 | 50	64 | 78
11	79 | 82	79 | 73	79 | 82
12	75 | 81	79 | 82	100 | 86
13	80 | 53	64 | 33	57 | 75
14	64 | 78	86 | 50	64 | 78
15	50 | 86	50 | 86	50 | 86
16	93 | 92	93 | 92	93 | 92
17	80 | 82	100 | 86	100 | 86
18	59 | 73	86 | 58	64 | 78
19	77 | 67	79 | 55	64 | 89
20	93 | 92	93 | 92	93 | 92
21	57 | 75	64 | 33	57 | 75
22	63 | 89	79 | 91	86 | 92
23	91 | 88	100 | 86	100 | 86
24	71 | 80	50 | 57	71 | 80
25	71 | 90	71 | 90	71 | 90
26	84 | 87	79 | 82	93 | 85
27	64 | 89	79 | 73	64 | 89
28	64 | 89	71 | 50	64 | 89
29	57 | 75	64 | 56	57 | 75
30	71 | 80	71 | 80	71 | 80
31	93 | 92	43 | 100	93 | 92
32	63 | 63	64 | 78	64 | 78
33	79 | 91	71 | 80	79 | 91
34	71 | 90	64 | 56	71 | 90
35	36 | 80	36 | 80	36 | 80
Total	74 | 82	73 | 72	75 | 84

TABLE VIII. THE PERCENTAGE OF CORRECT PREDICTED NODES PER VALIDATION CASE AND THE PERCENTAGE OF TRUE NEGATIVES CONTRIBUTING TO THE CORRECT PREDICTED NODES IN MODEL 2.

	The K2 algorithm	**Naïve Bayes**	**Frequency**
Case	***Correct per case (%) | True negatives (%)***		
1	95 | 78	84 | 75	89 | 82
2	84 | 75	58 | 82	79 | 80
3	74 | 64	63 | 50	58 | 91
4	74 | 79	42 | 88	84 | 81
5	68 | 77	63 | 75	74 | 79
6	79 | 80	79 | 67	79 | 80
7	74 | 79	42 | 38	74 | 79
8	100 | 79	79 | 80	95 | 83
9	79 | 80	79 | 73	84 | 81
10	74 | 71	68 | 54	68 | 77
11	63 | 100	84 | 69	79 | 80
12	95 | 78	68 | 77	95 | 83
13	53 | 100	47 | 33	84 | 81
14	68 | 77	74 | 57	68 | 77
15	63 | 83	63 | 75	79 | 80
16	84 | 100	74 | 86	95 | 83
17	79 | 73	95 | 78	68 | 77
18	53 | 100	53 | 60	74 | 79
19	74 | 64	74 | 57	63 | 83
20	68 | 92	79 | 93	84 | 94
21	63 | 75	63 | 58	68 | 77
22	84 | 94	63 | 92	84 | 94
23	79 | 80	84 | 81	100 | 84
24	63 | 100	42 | 50	79 | 80
25	68 | 92	68 | 62	68 | 92
26	79 | 100	47 | 78	95 | 83
27	63 | 92	63 | 50	63 | 92
28	63 | 83	63 | 50	68 | 85
29	63 | 75	63 | 58	68 | 77
30	74 | 79	79 | 80	79 | 80
31	95 | 83	47 | 89	89 | 88
32	68 | 77	68 | 69	74 | 79
33	68 | 92	53 | 80	74 | 93
34	58 | 91	63 | 58	63 | 92
35	37 | 86	79 | 33	42 | 88
Total	72 | 83	66 | 68	77 | 84

IV. DISCUSSION

In this paper, we made a comparison between two Bayesian approaches and a frequency approach on predicting offender output variables based on the input of crime scene and victim variables and investigated whether these approaches can be used as a decision tool to limit the incorporation of persons of interest into homicide investigations. Using a database of 233 solved Dutch single offender/single victim homicide cases we trained the three approaches in making the correct prediction. The trained K2 algorithm, Naïve Bayes and frequency approach were validated using 35 solved Dutch single offender/single victim homicide cases. Based on the results of this study two conclusions can be made: a) the non-intelligent frequency approach showed similar or better results than the intelligent Bayesian approaches and b) when the results on node level are analyzed one has to conclude that the usability of these approaches in practice, to limit the number of persons of interest to be incorporated into homicide investigations, should be questioned. In this section the implications of the results, the shortcomings of our research and opportunities for future research are outlined.

A. Comparison of approaches

The frequency approach showed similar or better results than the Bayesian approaches. That no differences were found between the K2 algorithm and frequency on the measures of OPA and CLA, is in line with the study of Baumgartner, Ferrari and Salfati [2]. The Naïve Bayes approach seems to perform less.

One possible explanation for this result is the great diversity of homicide cases. Homicide appears in many shapes and forms, ranging from criminal settlements to crime passionel, and from sexually motivated child murder to an argument gotten out of hand [12]. All these shapes and forms have different motives and perpetrators but share to some extent the same crime scene and victim characteristics. The possibility that perpetrators with different motives could leave similar crime scenes makes criminal profiling extremely difficult. The use of a database of discrete variables, leaving no room for interpretation, makes it possibly even harder to link crime scene and victim variables to characteristics of the perpetrator. This could help explain that frequency, where no input variables are used, showed the same or better results than the Bayesian approaches. These results are in favor of researchers that claim that criminal profiling lacks scientific support [11]. Whether the discrete character of the variables, leaving no room for interpretation, has influenced the results and a more case specific method, like behavioral case analysis [5], could assist criminal investigators by the incorporation and prioritization of persons of interest, should be the focus of future research.

A second explanation of the results could be the number of cases used to train the models. When data size decreases, the performance of algorithms can degrade dramatically [13]. Baumgartner, Ferrari and Palermo [1] conclude that to sufficiently train a database for criminal behavior, one needs a training database of thousand cases. Using learning curves of predictive accuracy and structural robustness they show that the K2 and K2' algorithms are stable and robust after thousand cases [1]. In this study 233 cases were used to train and 35 to validate the approaches. At the time of this research it was impossible to reach a greater number of cases. The Dutch police does not hold a database that contains detailed information on historical homicide cases. Due to the law on police data, limiting the storage of police data to a certain period after the case is closed, 268 solved single offender/single victim homicide cases were left to be analyzed. Future research should be conducted to explore whether the results change when more than thousand cases are used. When one believes that various categories of homicide differ to the extent that it could not be reduced to one model, these categories of homicide should have their own models (e.g. sexual motivated homicide), the number of homicide cases needed to make stable and robust models would be much greater. Future research should be conducted to explore whether the modelling of categories of homicide lead to better results.

B. The use as a decision tool

The usability of the three approaches to limit the incorporation of persons of interest, should be questioned. The OPA and the CLA do show promising results, but the analysis of the individual nodes show that true negatives contribute to a large extent to the predictive accuracy in all three approaches.

When the individual output variables are analyzed we see that true negatives contribute to a large extent to the percentage of correct predictions for many output nodes. The consequences of true negatives for the use of the approaches as a decision tool are outlined by an example. Two output variables of which the predictive accuracy depends heavily on true negatives are 'young offender between 17-21 years' and 'blood relative of the victim'. When these results are used to incorporate persons of interest, it would lead to the inclusion of everyone younger than seventeen and older than twenty-one or the inclusion of everyone except blood relatives of the victim. These examples show that the results do not meet the aim to limit the incorporation of persons of interest into homicide investigations. Only one true positive scored an accuracy over 80% in both models and for all three approaches and that was the prediction that the offender is 'male'. The incorporation of solely men will reduce the persons of interest considerably. In this case the question is whether a prediction with an accuracy around 80% is strong enough to exclude all women. When the results are used to prioritize between persons of interest that are already incorporated in the investigation, as proposed by Baumgartner, Ferrari and Palermo [1], one could prioritize men over women. Future research can be conducted to explore whether and how the results can be used to prioritize persons of interest.

When the percentage of correct predictions per validation case is analyzed the same dependency on true negatives stands out. The Naïve Bayes approach relied less on true negatives than the K2 algorithm and frequency approach. This seems an indication that Naïve Bayes is better able to predict true positives than the K2 algorithm and frequency approach. Future research can be conducted using other evaluation metrics to gain a better understanding of these differences. The dependency on true negatives by all three approaches, makes that the aim of this study to use the approaches as a decision tool to limit the incorporation of persons of interest into homicide investigations was not met. Future research should be conducted to investigate whether the adjustment of input and output variables can lead to models with better results. In order to investigate whether the used models are good in predicting output variables in specific categories of homicides, it will be interesting to conduct research into the similarities between validation cases in which a high or low percentage of nodes was predicted correct.

One extra complicating factor is the common omission of information in police data. The presence of unknowns makes, that in order to use Bayesian algorithms, choices have to be made on how to handle these unknowns. It is not clear whether Baumgartner, Ferrari and Salfati [2] used a database without unknowns or how they handled unknowns in their dataset. In this study we used two models that differ on how it handles the unknowns, leading to different models with different input and output variables and different results. Model 1, in which all variables holding unknowns were excluded, shows better overall results than Model 2, in which variables are included if there are less than 10% unknowns. Knowing that police data commonly holds unknown information, the exclusion of all variables holding unknowns is not a desired method. Future research has to be conducted to explore how to handle unknowns in a database of homicide investigations.

Sutmuller, den Hengst, Barros and van Gelder [12] showed that the perpetrator can be incorporated earlier into the investigation when categories are used to incorporate persons of interest into homicide investigations. However, by using these categories, a large number of persons of interest will be incorporated in the investigation [12]. This study aimed to limit these large numbers by using approaches that predict offender output variables based on the input of crime scene and victim variables. After the comparison of three approaches and the evaluation of the usability of those approaches, one has to conclude that the use of one general approach to predict offender characteristics based on crime scene and victim characteristics is not the solution to limit the incorporation of persons of interest into homicide investigations.

References

[1] K. Baumgartner, S. Ferrari, and G. Palermo, "Constructing Bayesian networks for criminal profiling from limited data," Knowledge-Based Systems, 21(7), 563-572, 2008.

[2] K. Baumgartner, S. Ferrari, and C.G. Salfati, "Bayesian network modeling of offender behavior for criminal profiling," Paper presented at the Proceedings of the 44th IEEE Conference on Decision and Control. p2702-2709, 2005.

[3] D.V. Canter, "Resolving the offender “profiling equations” and the emergence of an investigative psychology," Current Directions in Psychological Science, 20(1), 5-10, 2011.

[4] D.V. Canter, L.J. Alison, E. Alison, and N. Wentink, "The organized/disorganized typology of serial murder: Myth or model?," Psychology, Public Policy, and Law, 10(3), 293, 2004.

[5] H. Dern, C. Dern, A. Horn, and U. Horn, "The fire behind the smoke: A reply to Snook and colleagues," Criminal Justice and Behavior, 36(10), 1085-1090, 2009.

[6] C. Devery, "Criminal profiling and criminal investigation," Journal of Contemporary Criminal Justice, 26(4), 393-409, 2010.

[7] C. Dowden, C. Bennell, and S. Bloomfield, "Advances in offender profiling: A systematic review of the profiling literature published over the past three decades," Journal of Police and Criminal Psychology, 22(1), 44, 2007.

[8] A.M. Goodwill, R.J. Lehmann, E. Beauregard, and A. Andrei, "An action phase approach to offender profiling," Legal and Criminological Psychology, 21(2), 229-250. 2016.

[9] R.R. Hazelwood, and J.E. Douglas, "The lust murderer," FBI L. Enforcement Bull., 49, 18, 1980.

[10] D.A. Muller, "Qualitative approaches to criminal profiling as ways of reducing uncertainty in criminal investigations," Policing: A Journal of Policy and Practice, 5(1), 33-40, 2011.

[11] B. Snook, R.M. Cullen, C. Bennell, P.J. Taylor, and P. Gendreau, "The criminal profiling illusion: What's behind the smoke and mirrors?," Criminal Justice and Behavior, 35(10), 1257-1276, 2008.

[12] A.D. Sutmuller, M. den Hengst, A.I. Barros, and P. H. A. J. M. van Gelder, "Comparison of Methodologies Used in Homicide Investigations to Collect, Prioritize, and Eliminate Persons of Interest: A Case Study of Three Dutch Real-World Homicide Cases," Policing: A Journal of Policy and Practice, 2018.

[13] S. Yang, and K.C. Chang, "Comparison of score metrics for Bayesian network learning," IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems Vol. 3, pp. 2155-2160, 1996