Using Machine Learning to Detect ICT in Criminal Court Cases

Danique Sessink
University of Twente
PO Box 217, 7500 AE Enschede
The Netherlands

d.a.m.sessink@student.utwente.nl

ABSTRACT

ICT involvement in criminal court cases is not always explicitly mentioned, which results in less attention being paid to crimes involving ICT. This research aims to detect ICT involvement in criminal court cases and classify these cases based on certain features. Naïve Bayes is used as a machine learning algorithm to automatically classify the criminal court cases. As a result, this paper delivers a model that can detect ICT involvement in criminal court cases, the features which it is based on and the accuracy of this classification. A statistical analysis of the dataset is provided to show trends in ICT involved crimes.

Keywords

Criminal court cases, ICT involvement, text mining, machine learning, Naïve Bayes.

1. INTRODUCTION

Recently, an article was published in the news about how the registered level of crime in the Netherlands has decreased to that of 1980 [12][5]. Although the number of crimes has decreased in the Netherlands, the ratio between the different types of crime has shifted. Due to the growth of the Internet and other technologies in the past 20 years, crime involving information and communication technologies (ICT) has increased significantly. In 2016, 11% of all Dutch residents were victimized by cybercrime[1] [6]. Only 8% of the victims filed a police report.

In this paper, instead of using the term 'cybercrime', 'a crime involving ICT' is used to not only focus on the criminal court cases labeled as cybercrime, but to also be able to focus on criminal court cases where ICT played a role but which were not labeled as cybercrime. Not all crimes involving ICT are registered as such, as they may appear as computer-aided traditional crimes, or the involvement of ICT in the crime is ignored or the role of ICT in the crime is not explicitly mentioned [10][17]. Consequently, the number of crimes involving ICT may be much higher than originally thought and might become more relevant to fight and prevent. Therefore, it is interesting to investigate ICT involvement in crimes.

2. BACKGROUND

In this section, a background is given on the terminology in this research area. Furthermore, the format of the dataset that will be used is discussed as well as the use of machine learning.

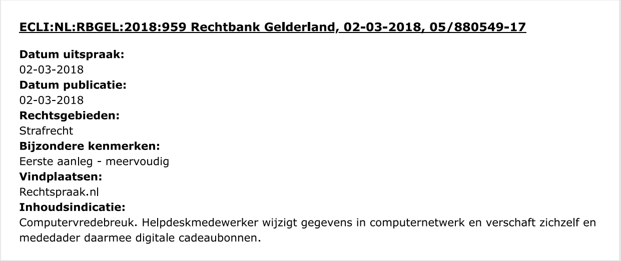
First, it is important to define ICT involvement in a crime and provide a clear overview of what aspects of ICT are included in this research. Choo et al. mention in their report a range of prefixes that are used to describe what they call technology-enabled crime, e.g. cyber-, ICT-related and digital [7]. In their report, the focus is on crimes that depend on ICT to make a profit, rather than crimes where ICT is merely a component. In this paper, all crimes with ICT (components) are considered, which is why the term 'ICT involvement' is used. Now, ICT involvement in a crime can be defined as the use of information and communication technologies before, during or after the crime. As 'information and communication technologies' is very broad, this term needs to be refined. In this research computers and the Internet are regarded as the main ICT components. Other ICT components could make use of these, e.g. accessing the Internet to post something with a smartphone, but the Internet will remain the main ICT component.

2.1 Traditional Crime vs. ICT involved Crime

It is difficult to draw the line between traditional crime and crime involving ICT. A strong difference between the two seems to be the geographical distance between the victim and the offender, as ICT allows for a greater distance between the two [11]. The boundary between traditional crime and ICT involved crime is constantly changing, especially with the increased use of the Internet [1]. Traditional crimes may have changed in the way they are executed, for example, some fraud crimes are now performed through the Internet which indicates a shift from traditional crime to ICT involved crime.

2.2 Format of the dataset

Since original police records of crimes cannot be obtained given the short time frame, court cases subject to criminal law are retrieved from the Internet. A short summary from such a criminal court case is shown in Figure 1. The full record of this criminal court case is publicly accessible online and contains much more information.



ECLI:NL:RBGEL:2018:959 Rechtbank Gelderland, 02-03-2018, 05/880549-17

Datum uitspraak:
02-03-2018
Datum publicatie:
02-03-2018
Rechtsgebieden:
Strafrecht
Bijzondere kenmerken:
Eerste aanleg - meervoudig
Vindplaatsen:
Rechtspraak.nl
Inhoudsindicatie:
Computervredebreuk. Helpdeskmedewerker wijzigt gegevens in computernetwerk en verschaft zichzelf en mededader daarmee digitale cadeaubonnen.

Figure 1. Brief example of a court case [15]

The obtained dataset consists of Dutch court cases subject to criminal law that are published on Rechtspraak.nl. A complete archive of all court cases from 1913 until 2018 can be

[1] CBS defines cybercrime as a form of crime which uses the Internet or other digital media to victimize people [6].

29thTwente Student Conference on IT, Jul. 6th, 2018, Enschede, The Netherlands. Copyright 2018, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

downloaded from this website, including non-criminal and non-published court cases. The case files will be saved in an .xml format.

The structure of a generic court case is not always the same. If a case from 2010 is compared with a case from 2018, the names of the sections are different and in the 2018 case the sections are numbered. If these results are compared with those of a case from 2016, the sections are different yet again and not numbered. However, the header of the cases is always presented in the same structure. In Table 1 the aspects of the header are briefly explained.

Table 1. Structure of header

Aspect	Content
Instantie	Which court the case was brought to
Datum uitspraak	The date of the judgement
Datum publicatie	The publication date of the case
Zaaknummer	Case number from the police
Rechtsgebieden	Judicial area(s)
Bijzondere kenmerken	Which level of court the case was brought to
Inhoudsindicatie	Summary of the case
Vindplaatsen	Places where the case was published

In addition, the names of the cases are structured. An European standard for numbering the cases is used, starting with 'ECLI', which stands for European Case Law Identifier [16]. The names of the cases are constructed like this: ECLI:countrycode:court code:year:number. An example of such a name is shown in Figure 1.

2.3 Machine Learning

With the use of machine learning, criminal court cases can be automatically classified based on certain features of ICT involvement which will be identified in this research.

From the number of cybercrimes that took place in 2016 and how much of those were reported to the police it can be concluded that 0.88% of all Dutch residents filed a police report for cybercrime. Domenie et al. support this number with their research, they conclude that the percentage of cybercrime in filed police reports is less than 1% [8]. Not all cases will go to court, so the percentage of cybercrime in criminal court cases will be even less. For training a classifier a large dataset is desirable. Since the size of the dataset was not yet determined and research has indicated the cybercrime rate in police reports is at most 1%, a provisionary choice was made for Naïve Bayes as the learning algorithm. The learning algorithm is effective and efficient for data mining [19] and proves to do well with little data [9][13].

From reading criminal court cases, certain classes were defined in which a case involving ICT could be classified. These categories can be found in Appendix A. Some categories have been removed. For example, if too little data was available for a category, it needed to be removed as more data was needed for correctly classifying files for this category. The remaining categories consist of: 'child pornography', 'cyberattack', 'identity theft', 'other', 'phishing', 'platform fraud' and 'online threat', with 'other' being a category a criminal court case will belong to if it does not fit into any of the defined categories.

3. RESEARCH QUESTIONS

In this research, to find out how many criminal court cases involve ICT, the aim is to answer the following questions:

RQ1 How can criminal court cases be classified as child pornography, cyberattack, identity theft, phishing or platform fraud based on ICT involvement?

- **RQ1.1** What features determine the classification of a criminal court case?
- **RQ1.2** Which model can be extracted from the classification of criminal court cases?
- **RQ1.3** What is the accuracy of the ICT involvement detection?

4. RELATED WORK

Some research is done on text mining and machine learning with crime detection, but not with criminal court cases as a dataset.

A master's thesis on text classification of Dutch police reports in which they try to find out if police reports can be classified through text mining, relates to this research topic but has made use of police reports instead of criminal court cases [4].

Androutsopoulos et al. compare in their research a Naïve Bayesian filter and a keyword-based filter, from which they conclude the viability of automatically trainable anti-spam filters [2]. In this research a Naïve Bayesian classifier is trained to automatically detect ICT involvement in criminal court cases.

A study done by Wang et al. on automatic document classification uses the Bayes' theorem as a basis for the algorithm to classify web documents [18]. One of the conclusions, which was consistent with earlier research, was that the multivariate Bernoulli event model performed worse than the multinomial event model classifier. This could be of interest for this research since it makes use of the Bayes' theorem for classifying the court cases.

There seem to be no studies that use the combination of Naïve Bayes as an algorithm and criminal court cases as a dataset, therefore the proposed research can create new insights.

5. METHOD OF RESEARCH

In this section, an overview of the method of research will be provided.

5.1 Data Collection

From Rechtspraak.nl, an archive containing all court cases from 1913 until 2018 was downloaded. The data was grouped per year and each year was divided into 12 folders, in which court cases were grouped per month. As the court cases were not grouped per district, e.g. criminal or civil, it was difficult to filter out the cases which were not required for this research. Fortunately, each file contains an identifier to indicate the district and another identifier to indicate if it concerns a published case or not. However, for filtering all the data every file needed to be accessed to check its identifiers, which is not a difficult process but could be a lengthy one when not automated.

Therefore, for filtering the data a Java program was executed to check the identifiers and relocate the unrequired files. The program also provided interesting statistics about the cases, for example how many cases exist per year, how many of them are published and how many of them belong to the criminal district.

The files before 1980 were deleted because they did not contain any published criminal court cases. From 1980 until April 2018, 2,406,921 cases exist. 443,217 of those cases are published and 83,889 of those are criminal court cases. The filtering process resulted in a usable dataset of 83,889 criminal court cases.

5.2 Naïve Bayes

For classifying the criminal court cases Multinomial Naïve Bayes is used, since probabilities need to be computed for several classes instead of one class. As a supervised learning setting will be used, a subset of the dataset will need to be labeled for the algorithm to predict the class of a court case based on what it learnt from the labeled data. The labeled subset can be split in train and test data to calculate the accuracy of the classifier.

5.3 Labeling

After obtaining the filtered dataset, court cases of the first two months of each year were read and labeled into the categories as listed in the table in Appendix A. If a court case did not belong to any of the pre-defined classes, it would be labeled as ‘other’. Categories could be removed or added during this process, depending on the number of cases in a certain category.

5.4 Classes

In total, 7 classes remained, including the ‘other’ class. The classes and number of files are shown in Table 2. The classes are imbalanced because each one of them does not contain the same number of files as the other. The imbalance of the classes will be taken into account when training the classifier.

Table 2. Classes and number of files

Class	Files
Child pornography	109
Cyberattack	51
Identity theft	39
Other	193
Phishing	51
Platform fraud	46
Online threat	51

5.5 Tools

For implementing the classifier, Python was used as a programming language since it provided useful toolkits for natural language processing. The toolkits that were used during the implementation were NLTK [3] and scikit-learn [14].

5.6 Pre-processing of the Data

As the files were downloaded in an .xml format, they needed to be stripped of the XML-tags first, which resulted in a text-only string. This string then needed to be loaded into a data frame, so the learning algorithm could process it. Before providing the classifier with the data from the data frame, a few actions needed to be performed on the data first.

5.6.1 Tokenize

From the NLTK toolkit, the `word_tokenize` function was used to divide a string into lists of substrings to get rid of any punctuation other than periods.

5.6.2 Stemmer

A Dutch Stemmer was used for stemming the dataset, which ensures words like ‘afbeelding’, ‘afbeeldingen’ and ‘afbeelden’ convert to the same word ‘afbeeld’. What the stemmer does is reduce the word to its base, which prevents words from the same meaning to be counted separately. The NLTK toolkit provides a stemmer for Dutch words.

5.6.3 Stop Words and Punctuation

Since the words the classifier will be based on are supposed to be meaningful, it is important to remove the stop words, punctuation and symbols from the text file. The NLTK toolkit includes a Dutch stop words set, which is convenient.

5.7 Training the Classifier

A classifier needs to be trained with the training set that is obtained from labeling many court cases. The top ten features can then be extracted from the dataset which help answer **RQ1.1**.

For training the classifier, a pipeline was created which contains both a vectorizer and the classifier itself. By creating a pipeline, all the logic can be put in one function call, which makes coding clearer and more error-proof. A `TfidfVectorizer()` was used to convert the files to a TF-IDF feature matrix. Subsequently, the `MultinomialNB()` function of scikit-learn was used to create the classifier. The attributes of both the `TfidfVectorizer` as well as the `MultinomialNB` classifier can be found in Table 3.

Table 3. Functions and attributes

Function	Attributes
`TfidfVectorizer()`	`norm=’l2’`
	`min_df=0.05`
	`max_df=0.75`
	`tokenizer=stemming_tokenizer`

	`stop_words=`
	`stopwords.words(‘dutch’) + list(string.punctuation) + symbols`

`MultinomialNB()`	`alpha=0.05`

The attributes `tokenizer` and `stop_words` are already explained in the previous section about the pre-processing of the data. An imbalance in the classes exists as mentioned before. The attribute `norm` in the `TfidfVectorizer` normalizes the term vectors counts, so the imbalance in the classes is no longer a problem. `min_df` and `max_df` are the thresholds for term frequency in a file. Terms that appear in less than 5% or more than 75% of the documents are ignored. The attribute `alpha` in the `MultinomialNB` function is the smoothing parameter, which is used to leave out noise.

Subsequently, the classifier was trained with the data from the data frame, which enabled for the top ten feature extraction per class. As a result, **RQ1.1** could be answered.

5.8 Extracting a model

A model was supposed to be extracted from the classification to answer **RQ1.2**. Since the Multinomial Naïve Bayes classifier is within the Python code and does not allow for a visualization of its model, another way of visualizing the model needed to be discovered. It could be interesting to investigate the probabilities of the features for the classes, and compare them to probabilities for that feature for other classes. Therefore, the probabilities for each class were calculated per feature, enabling a visualization of part of the Multinomial Naïve Bayes classifier.

5.9 Calculating the accuracy

Next, the accuracy of the classification needed to be determined to measure how well the algorithm performed on the dataset, answering **RQ1.3**.

For calculating the accuracy, `K-Fold` cross-validation was used. `K-Fold` cross-validation splits the dataset into `k` consecutive folds, from which each fold is used once as validation while the training set consists of the other folds. `K-Fold` cross-validation allows for calculating the accuracy of the classifier with the labeled dataset. In this classifier and label setting, 3 folds were used to do the cross-validation. The choice for this number is based on the low number of files per class. The `K-Fold` cross-validation enabled for calculating the `f1_score` for accuracy as well as creating a confusion matrix, which provides us more insight in the accuracy per class. The formula for the `f1_score` is as follows:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

In the formula for the `f1_score`, precision is:

$$Precision = \frac{true\ positives}{false\ positives + true\ positives}$$

And recall is:

$$Recall = \frac{true\ positives}{false\ negatives + true\ positives}$$

True positives and negatives, false positives and negatives can be put in a confusion matrix to show the performance of the classifier. An example of such a confusion matrix can be found in Table 4.

Table 4. Binary confusion matrix

Actual class	Predicted class		
		P	N
	P	True positive	False positive
	N	False negative	True negative

Table 4 shows a 2 by 2 confusion matrix for two classes, P and N. If the predicted class is P, and the actual class is P too, then it is a true positive, indicating the classification's prediction was right.

As multi-label classification is performed in this research instead of binary, the confusion matrix will consist of 7 rows and 7 columns, one row and one column per label. The confusion matrix is more detailed than the `f1_score`, as the `f1_score` in our case is the weighted average of the `f1_score` of every class. The confusion matrix shows how well the classifier performs for each class, including which classes are predicted as another class. For calculating the `f1_score`, the attribute `average='macro'` is used to calculate metrics for each class and find the unweighted mean for that class, as class imbalance is already taken into account.

Finally, the overall conclusions about detecting ICT involvement in criminal court cases could be drawn and will be discussed in the next sections.

6. RESULTS

In this section, the results of the research will be discussed.

6.1 Feature Extraction

The top ten features for every label were extracted per class. The results are shown in Table 5. The features make sense as these words are often associated with these sort of court cases.

Table 5. Top ten features per class

Class	Features
Child pornography	kinderporno, kennelijk, bezit, bestandsnam, afbeeld, kinderpornografisch, meisj, bereikt, seksuel, leeftijd
Cyberattack	wederrecht, val, benadeeld, computer, bedrijf, server, gegeven, werk, hof, geautomatiseerd
Identity theft	hof, bedrag, aangever, partij, belastingdienst, euro, benadeeld, nam, medeverdacht, slachtoffer
Other	getuig, cocain, arrest, gerechtshof, amsterdam, hoger, medeverdacht, beroep, slachtoffer, hof
Phishing	nam, bank, ing, phishing, geldbedrag, benadeeld, rekeninghouder, slachtoffer, organisatie, medeverdacht
Platform fraud	betal, oplicht, betrok, nam, bedrag, hof, euro, partij, benadeeld, slachtoffer
Online threat	nam, har, lev, geplaatst, tekst, bedreigd, bericht, hof, bedreig, slachtoffer

6.2 Model

A model would be extracted from the classifier to answer **RQ1.2**. This model can be found in Appendix B. It depicts feature prediction per class. The darker the value of the feature, the more probable that feature is for a class.

6.3 Confusion matrix and accuracy

The confusion matrix that was obtained from the classifier is depicted in Figure 2. It is in normalized form, since the classes are imbalanced. The darker the blue, the better the classifier is at predicting files for this class. It is clear where the classifier gets 'confused'. The 'identity theft' class does not seem to do well, which has a good reason. Through reading court cases, the discovery was made that 'platform fraud' is linked to 'identity theft', as it appears that stolen identities are often used to commit platform fraud. In the confusion matrix it is shown that 'identity theft' is often predicted as 'platform fraud'.

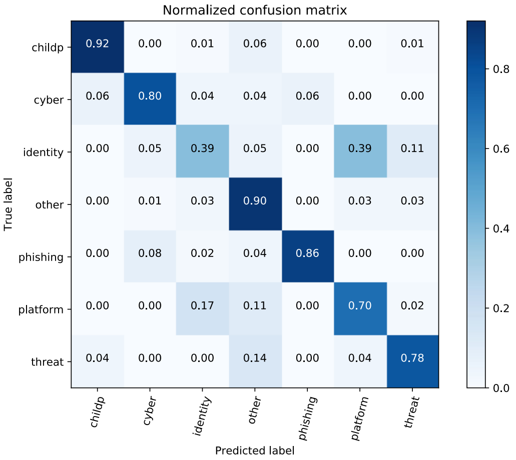


Figure 2. Normalized confusion matrix

From calculating the f1_score the accuracy proved to be 0.76, which means a criminal court case label can be predicted with an accuracy of 76%. This means 24% of all criminal court cases gets misclassified as another class. However, since this accuracy is the weighted average of each f1_score of a class, it may be better to calculate accuracies per class as some classes are performing better than others. The f1_score per class is shown in Table 6. The confusion matrix in Figure 2 clearly indicates as which classes the labels are misclassified, as well as the percentage per class. The accuracies can also be read from the diagonal in the confusion matrix. It appears 'child pornography' can be determined with high accuracy.

Table 6. Class accuracies

Class	Accuracy
Child pornography	92%
Cyberattack	80%
Identity theft	39%
Other	90%
Phishing	86%
Platform fraud	70%
Online threat	78%

Table 7. Code explanation

Code	Place
PHR	Parket bij Hoge Raad
RBMNE	Midden-Nederland
RBZUT	Zutphen
RBSGR	Groningen
RBROT	Rotterdam
RBAMS	Amsterdam
RBGEL	Gelderland
RBNNE	Noord-Nederland
HR	Hoge Raad
GHAMS	Amsterdam
RBUTR	Utrecht
GHLEE	Leeuwarden

6.4 Statistical analysis

It could be interesting to find out if certain trends exist among criminal court cases involving ICT. For example, how have these crime labels developed over the past years[2] and which courts do these cases usually go to. In Figure 3, the development of the defined crime labels is depicted.

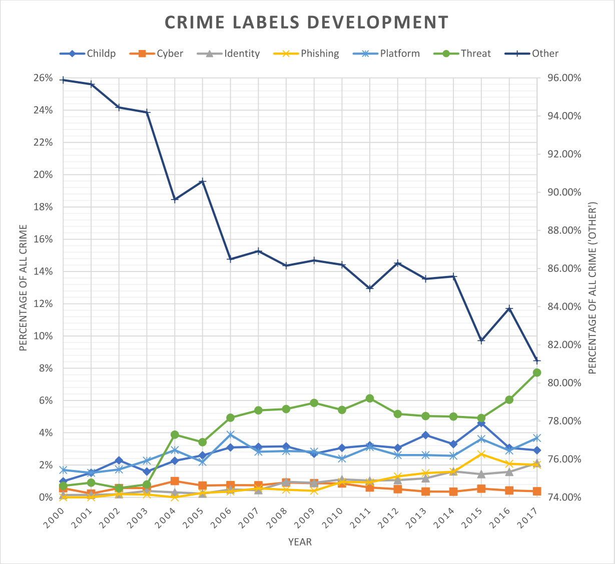


Figure 3. Crime development over the years

The crime labels are shown as percentage of total criminal court cases and are plotted against the years on the x-axis. The right y-axis values are only meant for the 'other' label, as this label is responsible for the largest percentage of criminal court cases. In the figure all the 'other' crime is decreasing while crime involving ICT seems to be increasing.

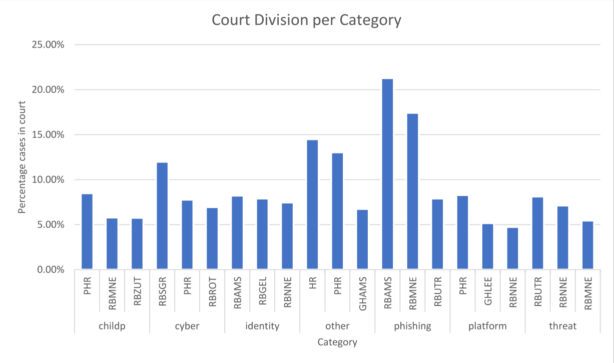


Figure 4. Top 3 courts per label

In Figure 4, the top 3 courts a case with a certain label usually goes to is shown, in percentage of the total criminal court cases within a label category. An explanation of the court codes can be found in Table 7. 'RB' before a code means 'Rechtbank' and 'GH' means 'Gerechtshof'. A criminal court case is initially covered by one of the 19 courts in the Netherlands. The court depends on the location where the crime (presumably) took place. Interesting is that 21,3% of all phishing cases are covered by 'RBAMS', meaning they are covered by the court of Amsterdam, but the crime of the case also seems to have taken place in Amsterdam. Since the place of the crime cannot always be determined in a phishing case, maybe the court of Amsterdam is taken as some sort of standard court where many of the phishing cases go to. It could also be that the case is covered in the region where the victim lives.

7. DISCUSSION AND FUTURE WORK

Throughout this research, decisions have been made that could influence the outcome of the research. This section will discuss on what ground these decisions have been made as well as the impact it could have on this research.

For labeling the data, the first two months of every year of data were read and subsequently labeled. This could introduce a bias, as there could be 'seasonal crimes' or less crime in the first two months of the year. Therefore, the dataset is not completely random, which is not preferable. However, since the categories are rather small, when completely randomly selecting the dataset and reading random files, many files would end up being classified as 'other', because it is the largest category. In the ideal situation all criminal court cases would need to be read and labeled, but this simply was not feasible given the time and the number of criminal court cases.

Often, a criminal court case includes more than one crime. For example, somebody who is guilty of committing phishing also breaks in to steal a TV, and both are treated in the same court case. Although the case would classify as 'phishing', if there are more cases like this, stealing TVs could also be regarded as 'phishing'. By reading child pornography cases, it was noticed that possession of child pornography usually is a 'byproduct' of fornication. So, the main case is then fornication, but the offender also possesses images that could classify as child pornography. Many cases like this exist, and in the end, it caused all fornication cases to classify as child pornography, even when they did not contain the child pornography part.

[2] The statistics depicted in Figure 3 and 4 rely on data between 2000 and 2017, because from 2000 and on a significant number of cases exist for the defined labels.

However, since about 250 cases were already labeled as 'child pornography', all cases with fornication were removed, meaning bias was reduced in the child pornography category. This was only done for this category. In the future, other categories could use the same bias reducing too.

Multinomial Naïve Bayes was used as the algorithm for classifying the criminal court cases. A master's thesis classifying Dutch police records indicates nice results with a SVM classifier [4]. However, due to time constraints such a classifier was never implemented during this research. In the future, this and other algorithms could be researched with the same sort of dataset to discover which algorithm performs best.

Regarding the size of the dataset, some of the categories were rather small, which could also introduce some bias. Most of that bias is already taken care of by the `TfidfVectorizer`, but a larger dataset is desirable (at least 100 files per category).

In the future, other categories could be introduced to find a variety of ICT related crimes.

8. CONCLUSION

Distinctive features exist that determine the classification of a criminal court case. These features usually consist of those associated with the crime, which makes sense. The top 10 features per class were extracted from the classifier.

A model that can be extracted from this classification is of course the Multinomial Naïve Bayes model, which does not allow for visualizing. However, visualizing the probability of a feature occurring in a class is possible, which was accomplished through a heat map.

The overall accuracy, calculated through the `f1_score` is 76%. Many differences exist between categories, which asked for a more detailed calculation of the accuracy, namely per class.

This research allows criminal court cases to be classified as 'child pornography', 'cyberattack', 'identity theft', 'other', 'phishing', 'platform fraud' and 'online threat'. In addition, in the future the classifier could possibly serve as a replacement for humans (e.g. court clerks) who label (ICT involved) court cases.

Moreover, the Dutch government and the police could look into the trends in ICT involved crimes discovered by this research and perhaps take measures to prevent such crimes.

9. ACKNOWLEDGEMENTS

I would like to thank my supervisors Marianne Junger and Erik Tews for guiding me throughout this research. Furthermore, I would like to thank Doina Bucur for her additional support.

10. REFERENCES

[1] R. Anderson, C. Barton, R. Böhme, R. Clayton, M. J. Van Eeten, M. Levi, T. Moore, and S. Savage. Measuring the Cost of Cybercrime. In *The economics of information security and privacy*, pages 265–300. Springer, 2013.

[2] I. Androutsopoulos, J. Koutsias, K. V. Chandrinos, and C. D. Spyropoulos. An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages. In *Proceedings of the 23rd annual international ACMSIGIR conference on Research and development in information retrieval*, pages 160–167. ACM, 2000.

[3] S. Bird, E. Klein, and E. Loper. Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media Inc., 2009.

[4] M. Brandenburg. Text Classification of Dutch police records. Master's thesis. Utrecht University, 2017.

[5] CBS. Het mysterie van de verdwenen criminaliteit. https://www.cbs.nl/nl-nl/achtergrond/2018/19/het-mysterie-van-de-verdwenen-criminaliteit [Online; accessed 9-May-2018]

[6] CBS. Veiligheidsmonitor 2016. https://www.cbs.nl/nl-nl/publicatie/2017/09/veiligheidsmonitor-2016. [Online; accessed 6-May-2018]

[7] K. K. R. Choo, R. G. Smith, and R. McCusker. *Future directions in technology-enabled crime: 2007-09*. Canberra: Australian Institute of Criminology, 2007.

[8] M. M. L. Domenie, E. R. Leukfeldt, M. H. Toutenhoofd-Visser, and W. Stol. Werkaanbod cybercrime bij de politie: een verkennend onderzoek naar de omvang van het geregistreerde werkaanbod cybercrime. *Lectoraat Cybersafety, Noordelijke Hogeschool Leeuwarden, Tech. Rep*, 23, 2009.

[9] G. Forman and I. Cohen. Learning from little: Comparison of classifiers given little training. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 161–172. Springer, 2004.

[10] P. H. Hartel, M. Junger, and R. J. Wieringa. Cyber-crime science= crime science+ information security. *CTIT, University of Twente, Technical Report TR-CTIT-10-34*, 2010.

[11] L. Montoya, M. Junger, and P. Hartel. How 'Digital' is Traditional Crime? In *2013 European Intelligence and Security Informatics Conference (EISIC)*, pages 31–37. IEEE, 2013.

[12] NOS. Criminaliteit in Nederland gedaald naar niveau van 1980. https://nos.nl/artikel/2230719-criminaliteit-in-nederland-gedaald-naar-niveau-van-1980.html [Online; accessed 9-May-2018]

[13] A. Y. Ng and M. I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In *Advances in neural information processing systems*, pages 841–848, 2002.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(Oct): 2825–2830, 2011.

[15] Rechtspraak.nl https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBGEL:2018:957 [Online; accessed 4-May-2018]

[16] Rechtspraak.nl https://www.rechtspraak.nl/Uitspraken-en-nieuws/Uitspraken/Paginas/ECLI.aspx [Online; accessed 12-May-18]

[17] D. S. Wall. *Cybercrime: The transformation of crime in the information age*, volume 4. Polity, 2007.

[18] Y. Wang, J. Hodges, and B. Tang. Classification of Web Documents Using a Naive Bayes method. In *Tools with Artificial Intelligence, 2003. Proceedings. 15thIEEE International Conference*, pages 560–564. IEEE, 2003.

[19] H. Zhang. The optimality of naive Bayes. In *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference*, pages 562–567. FLAIRS 2004, 2004.

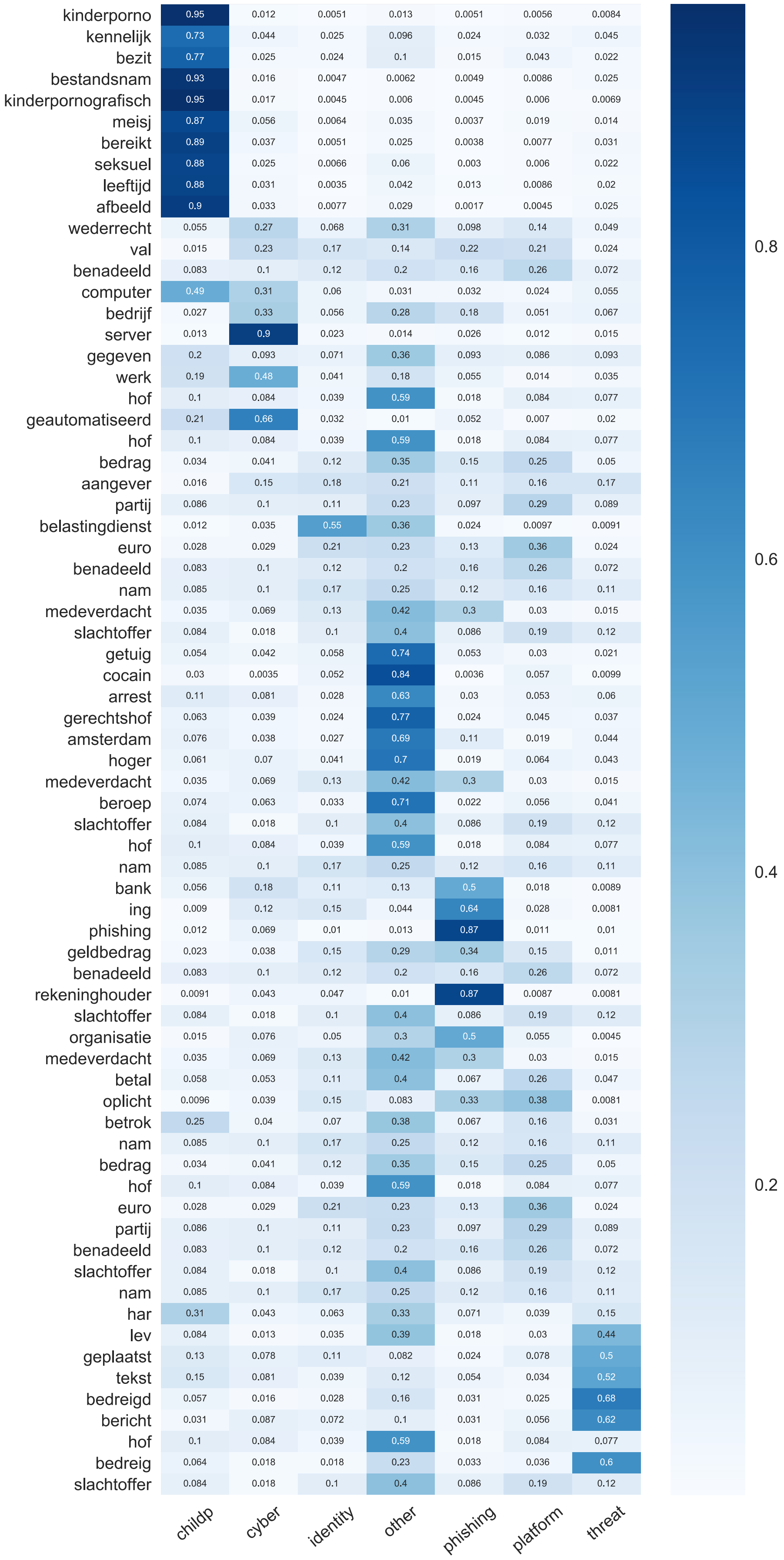
APPENDIX

A. TABLE OF CATEGORIES

Categories	Definition	Examples	Subcategories	Definition	Examples
Online deception	Using any form of digital means (e.g. a fake account) to deceive a person or organization for personal gain (e.g. money)	Using a fake account to buy something online but not paying for the product after receiving it	Phishing	Posing as a trustworthy entity through email or any other digital forms of communication to obtain sensitive information for personal gain	Send email to person to obtain credit card data to buy products with their money
			Online banking	Using banking data from other people or authority to obtain money from bank accounts through online means	Using logins from other people who trust you and transfer money from their bank account into your own bank account
			Identity theft	Using another individual's personal data through online means for gaining money	Opening a credit card in another person's name
			Tax fraud	Using a digital or online means to commit tax fraud	Filing a false tax return online
			Platform fraud	Using a platform (e.g. a website owned by offender or websites like 'Marktplaats[3]') to offer products as a seller and receive money, but never deliver to buyer	Offering a phone on a website, customer buys it, pays money before receiving the phone and seller never actually sends phone to buyer
Illegal content	Distribution of content that is forbidden to possess or distribute by law, by any digital means	Distribution of child pornography through the Internet	Child pornography	Storing or distributing child pornography by any digital means	Possess images of child pornography
			Copyright	Distributing any content which violates copyright	Publish copyrighted songs online without consent
Prohibited expression	Expression in words, media or other ways through any digital means that threatens or offends other people or groups of people	Publishing the contents of a book online that discriminates a group of people	Online threat	Express online that you want to harm a person or a group of people	Threaten to kill someone through a tweet on Twitter
			Online discrimination	Express online a prejudice of a certain person or group of people that is for example based on sex	Publish online that you think women are not equal to men
Violation of privacy	Using any form of digital means to violate a person's privacy which is prohibited by law	Capture videos of a person without their consent			
Cyberattack	Using any form of digital means to target an ICT system operated by either a person or an organization and steal, alter or destroy this system	Using a DDoS attack to overload a system to prevent users from using it for some time			

[3] Advertising website in the Netherlands

B. PREDICTED FEATURE RESULTS


kinderporno 0.95 0.012 0.0051 0.013 0.0051 0.0056 0.0084
kennelijk 0.73 0.044 0.025 0.096 0.024 0.032 0.045
bezit 0.77 0.025 0.024 0.1 0.015 0.043 0.022
bestandsnam 0.93 0.016 0.0047 0.0062 0.0049 0.0086 0.025
kinderpornografisch 0.95 0.017 0.0045 0.006 0.0045 0.006 0.0069
meisj 0.87 0.056 0.0064 0.035 0.0037 0.019 0.014
bereikt 0.89 0.037 0.0051 0.025 0.0038 0.0077 0.031
seksuel 0.88 0.025 0.0066 0.06 0.003 0.006 0.022
leeftijd 0.88 0.031 0.0035 0.042 0.013 0.0086 0.02
afbeeld 0.9 0.033 0.0077 0.029 0.0017 0.0045 0.025
wederrecht 0.055 0.27 0.068 0.31 0.098 0.14 0.049
val 0.015 0.23 0.17 0.14 0.22 0.21 0.024
benadeeld 0.083 0.1 0.12 0.2 0.16 0.26 0.072
computer 0.49 0.31 0.06 0.031 0.032 0.024 0.055
bedrijf 0.027 0.33 0.056 0.28 0.18 0.051 0.067
server 0.013 0.9 0.023 0.014 0.026 0.012 0.015
gegeven 0.2 0.093 0.071 0.36 0.093 0.086 0.093
werk 0.19 0.48 0.041 0.18 0.055 0.014 0.035
hof 0.1 0.084 0.039 0.59 0.018 0.084 0.077
geautomatiseerd 0.21 0.66 0.032 0.01 0.052 0.007 0.02
hof 0.1 0.084 0.039 0.59 0.018 0.084 0.077
bedrag 0.034 0.041 0.12 0.35 0.15 0.25 0.05
aangever 0.016 0.15 0.18 0.21 0.11 0.16 0.17
partij 0.086 0.1 0.11 0.23 0.097 0.29 0.089
belastingdienst 0.012 0.035 0.55 0.36 0.024 0.0097 0.0091
euro 0.028 0.029 0.21 0.23 0.13 0.36 0.024
benadeeld 0.083 0.1 0.12 0.2 0.16 0.26 0.072
nam 0.085 0.1 0.17 0.25 0.12 0.16 0.11
medeverdacht 0.035 0.069 0.13 0.42 0.3 0.03 0.015
slachtoffer 0.084 0.018 0.1 0.4 0.086 0.19 0.12
getuig 0.054 0.042 0.058 0.74 0.053 0.03 0.021
cocain 0.03 0.0035 0.052 0.84 0.0036 0.057 0.0099
arrest 0.11 0.081 0.028 0.63 0.03 0.053 0.06
gerechtshof 0.063 0.039 0.024 0.77 0.024 0.045 0.037
amsterdam 0.076 0.038 0.027 0.69 0.11 0.019 0.044
hoger 0.061 0.07 0.041 0.7 0.019 0.064 0.043
medeverdacht 0.035 0.069 0.13 0.42 0.3 0.03 0.015
beroep 0.074 0.063 0.033 0.71 0.022 0.056 0.041
slachtoffer 0.084 0.018 0.1 0.4 0.086 0.19 0.12
hof 0.1 0.084 0.039 0.59 0.018 0.084 0.077
nam 0.085 0.1 0.17 0.25 0.12 0.16 0.11
bank 0.056 0.18 0.11 0.13 0.5 0.018 0.0089
ing 0.009 0.12 0.15 0.044 0.64 0.028 0.0081
phishing 0.012 0.069 0.01 0.013 0.87 0.011 0.01
geldbedrag 0.023 0.038 0.15 0.29 0.34 0.15 0.011
benadeeld 0.083 0.1 0.12 0.2 0.16 0.26 0.072
rekeninghouder 0.0091 0.043 0.047 0.01 0.87 0.0087 0.0081
slachtoffer 0.084 0.018 0.1 0.4 0.086 0.19 0.12
organisatie 0.015 0.076 0.05 0.3 0.5 0.055 0.0045
medeverdacht 0.035 0.069 0.13 0.42 0.3 0.03 0.015
betal 0.058 0.053 0.11 0.4 0.067 0.26 0.047
oplicht 0.0096 0.039 0.15 0.083 0.33 0.38 0.0081
betrok 0.25 0.04 0.07 0.38 0.067 0.16 0.031
nam 0.085 0.1 0.17 0.25 0.12 0.16 0.11
bedrag 0.034 0.041 0.12 0.35 0.15 0.25 0.05
hof 0.1 0.084 0.039 0.59 0.018 0.084 0.077
euro 0.028 0.029 0.21 0.23 0.13 0.36 0.024
partij 0.086 0.1 0.11 0.23 0.097 0.29 0.089
benadeeld 0.083 0.1 0.12 0.2 0.16 0.26 0.072
slachtoffer 0.084 0.018 0.1 0.4 0.086 0.19 0.12
nam 0.085 0.1 0.17 0.25 0.12 0.16 0.11
har 0.31 0.043 0.063 0.33 0.071 0.039 0.15
lev 0.084 0.013 0.035 0.39 0.018 0.03 0.44
geplaatst 0.13 0.078 0.11 0.082 0.024 0.078 0.5
tekst 0.15 0.081 0.039 0.12 0.054 0.034 0.52
bedreigd 0.057 0.016 0.028 0.16 0.031 0.025 0.68
bericht 0.031 0.087 0.072 0.1 0.031 0.056 0.62
hof 0.1 0.084 0.039 0.59 0.018 0.084 0.077
bedreig 0.064 0.018 0.018 0.23 0.033 0.036 0.6
slachtoffer 0.084 0.018 0.1 0.4 0.086 0.19 0.12
childp
cyber
identity
other
phishing
platform
threat