Cri-Astrologer: Predicting Demography of Involved Criminals based on Historical Data

Md. Atiqur Rahman
A. B. M. Alim Al Islam
atiqur.rahman0041@gmail.com
razi_bd@yahoo.com
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

ABSTRACT

Because of the rapid advancement in computer technology, police enforcement agencies are now able to keep enormous databases that contain specific information about crimes. These databases can be utilized to analyze crime patterns, criminal characteristics, and the demographics of both criminals and victims. Through the application of various machine learning algorithms to these datasets, it is possible to generate decision-aid systems that can assist in the conduct of police investigations. When there is a large amount of data accessible, several data-driven deep learning approaches can also be utilized. Within the scope of this investigation, our primary objective is to create a tool that may be utilized during the standard investigative process. To forecast criminal demographic profiles using crime evidence data and victim demographics, we present a deep factorization machine-based DNN architecture. We evaluate the performance of our architecture in comparison to that of traditional machine learning algorithms and deep learning algorithms, and we provide our findings in a comparative study.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; • **Applied computing** → *Investigation techniques*.

KEYWORDS

datasets, deep learning, criminal profiling

ACM Reference Format:
Md. Atiqur Rahman and A. B. M. Alim Al Islam. 2022. Cri-Astrologer: Predicting Demography of Involved Criminals based on Historical Data. In *2022 9th International Conference on Networking, Systems and Security (NSysS 2022), December 20–22, 2022, Cox's Bazar, Bangladesh.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3569551.3569561

1 INTRODUCTION

Criminal profiling is the process of determining personality traits, behavioral inclinations, geographic locations, and demographic or biological descriptions of an offender based on the factors of the crime [14]. It has been practiced for a long time by law-enforcement investigators. Profilers' processes to analyze crime factors are either Clinical or Statistical [30]. Clinically oriented techniques incorporate aspects of the profilers' intuition, knowledge, experience, and training to generate predictions where statistically oriented predictions are based upon descriptive and inferential statistical models derived from an analysis of traits of offenders who have committed similar types of crime before [30]. Many do not have faith in human profilers because of deficiencies in the empirical literature regarding the exactness of criminal profiling procedures. In our study, we try to explore how machine learning and deep learning algorithms perform in this sector, given different crime factors.

1.1 Motivation behind Our Work

Although offender profiling has been practiced for a long time, many investigators do not believe it helps in the investigation procedure. It is partly because of the human factor present in the procedure. Humans can be biased towards or against different factors and make wrong judgments. Such actions can derail the investigation procedure even, sometimes resulting in the wrong person getting punishment [7]. It is best to remove the human factor to avoid the aforementioned scenarios. Machine learning does that, so we explore the machine learning approach in criminal profiling. Considering the recent advancement of deep learning in research and development, we also explore different deep learning techniques in this sector.

1.2 Limitations of The Existing Studies

Criminal profiling using machine learning and data analysis is not typical in the literature. There are numerous researches on crime prediction and prevention, but what follows a crime? Existing research on criminal profiling does not concentrate on forecasting criminal demography from a general perspective. In addition, they do not investigate big data and deep learning techniques. We attempt to gain a broader understanding of criminal profiling and utilize massive datasets.

1.3 Research Questions

Criminal profiling has historically been a field dominated by human profilers. In this industry, statistical analysis and machine learning techniques are uncommon. Given the limitations of human profilers, few have faith in this technique. Some do not even consider it a scientific approach due to the lack of evidence [30]. We intend to introduce the concept of deep learning to this field in order to

NSysS 2022, December 20–22, 2022, Cox's Bazar, Bangladesh

ACM ISBN 978-1-4503-9903-6/22/12…$15.00
https://doi.org/10.1145/3569551.3569561

exploit the wealth of data available today. Based on the above discussions, we have three research questions for this study:

RQ1: *Can we predict a criminal's demographic profile based on historical crime incidence records?*

We investigate various machine learning methods and popular and open-source deep learning algorithms to predict the criminal demographic features we desire to answer this question. We compare their performances and evaluate the resulting data.

RQ2: *Is it possible to overcome the limitations of human practitioners in generalized criminal demographic profiling with machine learning techniques?*

We make it a priority to instill non-discriminatory attitudes in our models, whether it be with regard to gender, race, ethnicity, or nationality.

RQ3: *Can we develop a novel DNN architecture that can perform better than existing architectures in predicting criminal demographic profiles in a generalized manner?*

We try to develop a DNN architecture that will have a faster and more efficient performance compared to the ones that already exist.

1.4 Our Contributions

In addition to the conventional methods of conducting a research study, such as examining the datasets and employing conventional machine learning and deep learning algorithms, we attempted to do some novel research.

- We present a new DNN architecture that can predict criminal demographics from crime evidence data and victim demographics. We design the DNN based on various open-source implementations of deep learning techniques taken from the relevant research literature.
- We use two real-life crime incident-based datasets never before used for criminal demographic profiling. We explore their complex structure and attempt to make it more understandable. We apply classical machine learning and deep learning algorithms to these datasets for predicting criminal demographics.

1.5 Organization of this Study

In this study, we describe our overall methodology in great depth, including how we carry out the tests, the models that we employ, and the reasoning behind our selection of those models. After that, we show the experimental results we acquire from the various models and present an analysis that contrasts the models. In conclusion, we explore the limits of this study and the potential future work that we may perform in light of our findings.

2 RELATED WORK

Criminal profiling has been used by investigators for a long time and they use different strategical approaches. With increasing data resources about crime data and corresponding offender demographic, it seems feasible to approach statistical or machine learning approach. We reviewed different literature in this study. We divide them into three subsections and discuss their approaches and limitations.

2.1 Work on Criminal Profiling

The study conducted by South & Messner [31] explore the impact of age, sex, and race on criminal behavior. It also contrasts between compositional and contextual effects of demographic structure on aggregate crime rates. Snook et al., [30] conduct a systematic review of the literature to determine the scientific credibility of criminal profiling. They try to inform the utility of this specific investigative technique for practitioners. Devery [7] discusses the instability of the traditional criminal profile system and raises the question of whether criminal profiling should be given importance in an investigation or not. He points out the pitfalls of criminal profiling by human practitioners. On the contrary, Cook & Hinman [6] show an optimistic point of view on criminal profiling. They point out the differences between the fiction and reality of criminal profiling. They also point out its scientific basis. Rich [27] presents a study about applying machine learning algorithms to government data to identify probable crime suspects and thus prevent crime. He describes various Automated Suspicion Algorithms in his study, which can identify data-supported correlations between innocent behavior and criminal activity. Zheng et al., [35] propose a passenger profiling method for airlines based on fuzzy deep machine learning. They develop a deep neural network for classifying ordinary passengers and potential attackers and further develop an integrated deep neural network for identifying group attackers whose individual features are insufficient to reveal the abnormality. Baumgartner et al., [5] propose a bayesian network model of offender behavior by analyzing the action of an offender at the crime scene to his behavioral profile. They use structural and parameter learning algorithms to discover inherent relationships embedded in a database containing crime scene and offender characteristics from homicide cases solved by the British police from the 1970s to the early 1990s. They use a technique to reduce the search space of possible Bayes-net structures by modifying the greedy search K2 learning algorithm to include apriori conditional independence relations among nodes and an inference algorithm to predict the offender profile from the behaviors observed on the crime scene. They show around 15% improvement from a model obtained from the same data by the original K2 algorithm.

2.2 Work on National Incident-Based Reporting System (NIBRS) Data

Akiyama & Nolan [2] discuss the complexity of NIBRS data. They provide an overview of the NIBRS structure and methods for maneuvering within it to present and interpret cross-tabulations of the NIBRS data correctly. Addington [1] discusses current examples of how criminologists use the UCR data and issues to consider when working with fully incident-based UCR, especially concerns not present with aggregate crime data. Krienert et al., [15] present a study on five years of National Incident-Based Reporting System (NIBRS) data (2008–2012) that provide baseline information on reported male-to-female marital sexual intimate partner violence (IPV) compared to nonmarital sexual IPV. They found that husbands as offenders and wives as victims are significantly older than non-married sexual IPV offenders and victims. Married offenders are more likely to be White, and dating offenders are Black. Injuries are more likely if the victim and offender are married. Marital sexual

IPV cases are more likely to include sexual penetration, including a higher incidence of rape, sodomy, and sexual assault with an instrument. Lamari et al., [16] present an efficient machine learning framework that can predict spatial crime occurrences without using past crime as a predictor and at a relatively high resolution: the U.S. Census Block Group level. They propose a framework based on an in-depth multidisciplinary literature review allowing the selection of 188 best-fit crime predictors from NIBRS data. They select the predictive models by conducting a comparative study of different machine learning algorithms, including generalized linear models, ensemble learning, and deep learning. The gradient boosting predicts most accurately for violent crimes, property crimes, motor vehicle thefts, vandalism, and the total crime count. Their proposed framework achieves 73% and 77% accuracy while predicting property crimes and violent crimes, respectively.

2.3 Work on NYPD Complaint Data

Mehranbod et al., [18] present a study about whether ridesharing is associated with an increased incidence of alcohol-related assaults in New York. Their research concludes that ridesharing is positively associated with nighttime assaults at bars but not at restaurants. Almuhanna et al., [3] propose a methodology to predict Spatio-temporal criminal patterns within New York City neighborhoods using different machine learning classifiers. XGboost predicts the highest number of correct classifications out of 25 different crime types. It accurately predicts 22 types of crime, whereas Random Forest predicts 21 types of crime, and Support Vector Machine predicts 17 types of crimes with the lowest accuracy.

2.4 Gap in the Literature

The literature review above shows that Criminal profiling with machine learning and data analysis is not standard in the literature. In the study conducted by Rich [27], he proposes a machine learning approach to prevent crime but not to profile criminals after a crime incident has happened. Zheng et al., [35] propose a passenger profiling method for airlines, which is quite specific and does not generalize the overall criminal profiling sector. Baumgartner et al., [5] propose a bayesian network approach to predict offenders' behavioral profiles from limited data. They do not work with extensive crime incident databases.

From the datasets literature review, we can see that there has been almost no work in the literature on these databases for criminal profiling. Krienert et al., [15] give a statistical analysis of Intimate Partner Violence. They try to statistically present some demographic features of the offenders for a specific crime type. Lamari et al., [16] propose a machine learning framework to predict spatial crime occurrence, but that also does not involve criminal profiling. Mehranbod et al., [18], and Almuhanna [3] work with NYPD complaint data, but they do not explore this sector also. We try to fill this gap in the literature by working with these two datasets for offender demographic profiling.

3 BACKGROUND AND PRELIMINARIES

3.1 Chi-Square Test

The Chi-square test is used for two types of statistical analysis: the test of independence and the test of goodness of fit. In feature selection, it is used to test whether the class label is independent of a feature. Chi-Square test score with C class and r values is defined as [8, 25]:

$$\chi^2 = -\sum_{\mathrm{i}=1}^{\mathrm{r}} \sum_{\mathrm{j}=1}^{\mathrm{C}} \frac{\mathrm{n_{ij}} - \mu_{\mathrm{ij}}}{\mu_{\mathrm{ij}}} \tag{1}$$

where $\mathrm{n_{ij}}$ is the number of samples value with the $\mathrm{i^{th}}$ value of the feature. μ_{ij} is defined as [8, 25]:

$$\mu_{\mathrm{ij}} = \frac{\mathrm{n}_{*\mathrm{j}} \times \mathrm{n}_{\mathrm{i}*}}{\mathrm{n}} \tag{2}$$

Where $\mathrm{n}_{\mathrm{i}*}$ is the number of samples with the $\mathrm{i^{th}}$ feature value, $\mathrm{n}_{*\mathrm{j}}$ is the number of samples in class j, and n is the total number of samples. We aim to find features on which class label is highly dependent.

3.2 One-Hot Encoding

In one-hot encoding, we create a new variable for each class of a categorical feature. We map each class with a binary variable containing either 0 or 1. Here, 1 represents the presence, and 0 represents the absence of that class. We call these newly created binary features dummy variables. The number of dummy variables equals the number of classes present in the categorical feature. We use the one-hot encoding technique when the features are nominal (do not have any order) [28, 29]. For example, Suppose we have a dataset with a categorical feature named Weapon Used, having different classes like Gun, Knife, Machete, and Hand. Now we one-hot encode this data which is shown in Figure 1.

index	Weapon Used
0	Gun
1	Machete
2	Knife
3	Hand

index	Gun	Machete	Knife	Hand
0	1	0	0	0
1	0	1	0	0
2	0	0	1	0
3	0	0	0	1

Figure 1: One-Hot encoding

After encoding, we have dummy variables representing a class in the feature Weapon Used in the second table. Now for each class present, we have 1 in the column of that category and 0 for the others.

3.3 Label Encoding

In Label encoding, we convert each categorical class into an integer value. We use this categorical data encoding technique when the categorical feature is ordinal. In this case, maintaining order is essential. Hence encoding should reflect the sequence. We also

Index	Weapon Used
0	Gun
1	Machete
2	Knife
3	Hand
4	Gun

Index	Weapon Used
0	1
1	3
2	2
3	4
4	1

Figure 2: Label encoding

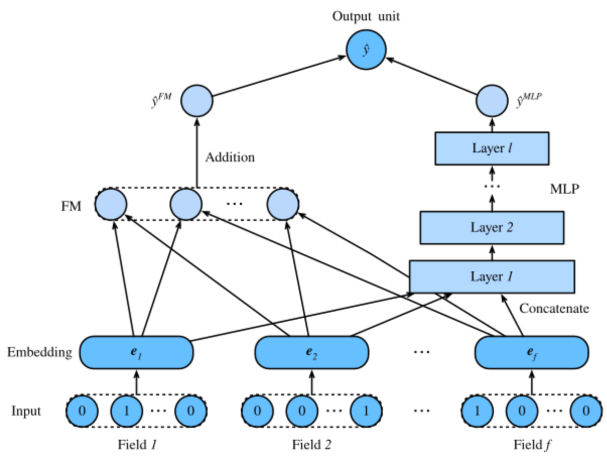

Figure 3: DeepFM architecture [9]

use this encoding technique to encode class labels [28, 29]. For the categorical feature example given in Subsection 3.2, we label-encode the data which is shown in Figure 2.

3.4 Deep Factorization Machine

A Factorization Machine (FM) component and a deep component are merged into a parallel structure to form a Deep Factorization Machine (DeepFM). The FM component is similar to the two-way factorization machines used to describe low-order feature interactions. A Multi-Layer Perceptron (MLP) captures the deep component's high-order feature interactions and nonlinearities. The inputs/embeddings for these two components are the same, and their outputs are added together to get the final forecast. DeepFM's concept is similar to the Wide and Deep architecture, which can capture both memorization and generalization. DeepFM, as opposed to the Wide and Deep models, saves time and effort by automatically identifying feature combinations [9].

4 METHODOLOGY OF OUR STUDY

This chapter elaborately discusses our methodology, explaining the different approaches we have explored. Before going into detail, we present an overview of our methodology in the next section.

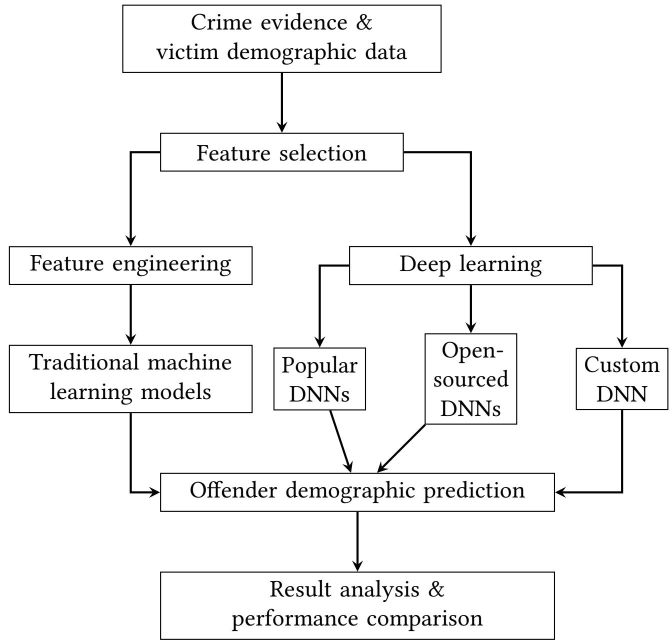

Figure 4: Flow chart on the methodology of our study

4.1 Overview of Our Methodology

In this study, we explore two types of data to predict offender demographic: *Crime Evidence Data* and *Victim Demographic Data*. This type of data usually contains many features which are not relevant to our research topic. So, we have to go through feature selection to reduce the dimensionality of our dataset and keep the most relevant features needed to predict offender demographic. After that, we use different feature engineering techniques to make the data usable for machine learning models. We use the transformed data and try different machine learning models to check their performances. Because our data volume is high, we also try a deep learning approach. We explore some existing deep learning models. We also use open-source deep learning models which work with data similar to ours. Finally, we use a custom deep learning model to check further if we can improve the prediction accuracy. We compare the performances of all the systems we have experimented and present a comparative analysis.

4.2 Data Collection

We explore two publicly available datasets in our study. Since our study's purpose is to predict offender demographic from crime evidence and victim demographic, we looked for crime incident record-based datasets that contain all of the required information for this study.

4.2.1 National Incident-Based Reporting System, 2016. The National Incident-Based Reporting System (NIBRS) is an incident-based reporting system utilized by United States law enforcement agencies to collect and report data on crimes. Local, state, and federal agencies create NIBRS data from their records management systems. [33] Incident-based data provide a substantial amount of information about crimes. The data is also structured in a sophisticated manner, representing the many distinct components of a crime incident [20]. A range of data is collected regarding each crime incidence that comes to the notice of law enforcement. This information includes the specific nature and types of offenses committed during the incident, the characteristics of the victim(s) and offender(s), the types and value of items stolen and recovered, and the characteristics of those arrested in connection with a crime incidence [19]. Every occurrence and arrest in the Group A offense category is recorded. There are 52 offenses in Group A, divided into 23 crime categories. The details of these offenses are compiled and reported to the National Institute of Justice. In addition to the Group A offenses, 10 Group B offenses with simply the arrest details are recorded [33].

In this dataset, there are four types of data files. Among them, we use Incident-Level Extract File. It contains one record for every NIBRS crime incident with an incident date in 2016. This file's overall number of records is 5,293,536, containing 390 variables. Other NIBRS variables from the offense, property, victim, offender, and arrestee segments were combined with the incident records using the ORI and INCIDENT NUMBER variables. Variables from the Batch Header segment were combined using the ORI variable. Records from the Window Exceptional Clearance segment were added to the file, given that the incident date was in 2016 [19].

4.2.2 NYPD Complaint Data Historic. The NYPD complaint data includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) from 2006 to the end of 2019 [23]. The New York City Police Department keeps track of reported crime and offense data under the New York State Penal Law. The information is not organized in the FBI's Uniform Crime Report (UCR) format and is not directly comparable. The FBI reorganizes the New York State Penal Law categories to provide national statistics that are comparable across all states' penal codes. Nevertheless, the data is categorized and evaluated similarly to the UCR. The reported instances are first categorized to identify all possible offenses, then scored to determine the most severe offense. Attempts to commit a crime are classified as a crime. All crime and offense complaint data based on New York State laws for incidents occurring within the borders of New York City is included. The list excludes federal criminal charges [21]. The dataset contains 7.38 million rows and 35 columns. Each row is a complaint. It was made public on 16 November 2016, and it is updated annually [23].

4.3 Feature Selection

Feature selection reduces the number of input features when creating a predictive model based on the redundancy and relevance of the data. It is proved from the literature that feature selection can improve the performance of prediction, scalability, and generalization capability of the classifier, and it plays a fundamental role in reducing computational complexity, storage, and cost [32].

We mainly use the Chi-Square method described in subsection 3.1 to select K-best features from the entire dataset. We experiment with different values for K, for example, K = 13, 25, 50. We cross-check the feature set to see whether any vital feature is left out.

Gun	Knife	Hand	No Weapon
1	1	0	0

Figure 5: Handling array type data with One-Hot encoding

Gun	Knife	Hand	No Weapon
0	0	0	0

Figure 6: Handling missing values with One-Hot encoding

4.4 Feature Engineering

Feature Engineering is a machine learning technique that uses data to generate new features never present in the original dataset. It can produce new features which can simplify and speed up data transformations as well as enhance model performance [22]. In the feature engineering process, a model's feature vector is expanded by adding new features that are calculated based on the other features [10]. Feature engineering consists of various processes. We apply two of those in our study: *Feature Creation* and *Feature Transformation.*

4.4.1 Feature Creation. Creating features involves identifying the features that will be most useful in the predictive model. This process requires human intervention and creativity. Existing features are mixed via addition, subtraction, multiplication, and ratio to create new derived features with greater predictive power. For the dataset in Subsubsection 4.2.1, we computed *Total Population of An Area* adding five features.

4.4.2 Feature Transformation. Feature transformation is a function that transforms features from one representation to another. In our study, we used some feature transformation techniques. We transformed *Date of Incident* into a categorical feature containing the month's name. We transformed *Time of Incident* into a categorical feature containing three classes, e.g., morning, evening, and night. We transformed the numerical feature *Age* to categorical feature *Age Range* because age range is more relevant to our study than specific age value. We used One-Hot Encoding described in subsection 3.2 to transform all the categorical input features because machine learning models need numerical data to operate properly. We used Label Encoding described in subsection 3.3 to transform class labels.

We introduce some changes in the one-hot encoding method based on our dataset characteristics. There are scenarios where multiple weapons are used in a crime incident, or an offender might be charged with multiple felonies. We set 1 to all the dummy features present in the scenario to incorporate these scenarios. Suppose *Weapon Used* feature has four classes: Gun, Knife, Hand, No Weapon. Then, we encode an incident with both gun and knife shown as in Figure 5. We also handle missing values with all 0's. Given an incident record with missing *Weapon Used* information, we get the encoding shown as in Figure 6. We drop all the records with missing criminal demographic information from our training dataset because those features are compulsory to train the model. Because one-hot encoding increases the size of the dataset, we also explore feature embedding in the deep learning approach.

5 APPLICATION OF TRADITIONAL MACHINE LEARNING MODELS

After completing the preprocessing steps, we fit our data to some popular machine learning models and check their performance. The commonly used machine learning algorithms are Linear Regression, Logistic Regression, Decision Tree, Support Vector Machine, Naive Bayes, k Nearest Neighbor, K-Means Clustering, Random Forest, and Gradient Boosting algorithms [26]. Because our data is fully categorical and nominal, and we are trying to solve a classification problem, we use models that can work with this type of data. We use *Logistic Regressionl*, *Decision Tree*, *Random Forest*, and *Naive Bayes* models in this study.

6 DEEP LEARNING APPROACHES

The datasets we work with are massive in size. We also explore deep learning approaches to see whether we could acquire better results. We explore some established deep learning models and open-source deep learning models, and finally, we propose a custom deep learning model and compare their performances.

6.1 Popular DNNs

We use some popular deep learning algorithms to check how they perform with our data. We apply 1-D ConvNet and Stacked RNN to our data and compare their performances with the machine learning models. We use the 1-D ConvNet structure from Kamal et al., [13], and the Stacked Bidirectional LSTM structure from Althelaya et al., [4] in this study.

6.2 Open Source DNNs

We explore existing literature to look for open-sourced deep learning works to incorporate into our study. We use the study conducted by Zhang et al., [34] where they deal with predicting click-through rates (CTR) given certain web features. The difference from continuous raw features that we usually find in the image and audio domains is that the input features in web space are always multi-field and are mostly discrete and categorical, while their dependencies are little known. This scenario is quite similar to our data. So, we can incorporate the methods used here and apply them to our dataset. This paper proposed two DNN architectures: Factorization-machine supported Neural Networks (FNN) and Sampling-based Neural Networks (SNN). The categorical input features are field-wise one-hot encoded. They evaluated their models based on the iPinYou dataset [17].

We also explore an upgraded version of the work mentioned above. Qu et al., [24] worked with a Product-based Neural Network. It consists of an embedding layer to learn a sparse representation of the categorical data, a product layer to capture interactive patterns between inter-field categories, and further fully connected layers to explore high-order feature interactions. We use their models to study how they work with our data. Since their code used a bit of backdated modules, we reworked their code to some extent. They addressed a binary classification problem with these models. So we had to modify our multiclass classification problem to a multiple binary classification problem and redesign our data as required.

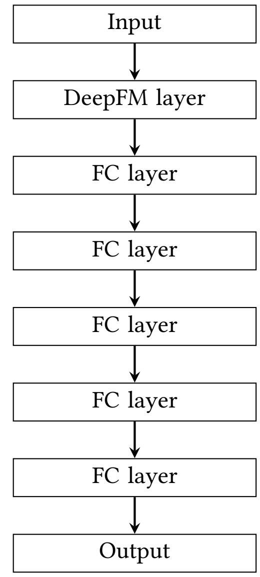

Figure 7: Custom DNN architecture

6.3 Custom DNN

We propose a custom deep neural network for predicting criminal demographics from crime evidence data and victim demographics. We use a modified DeepFM layer described in Section 3.4 followed by multiple fully connected layers, each having *ReLU* as an activation function. We design each fully connected layer having neurons in decreasing 2's exponential. In the final layer, we use *Softmax* activation.

7 RESULT ANALYSIS AND PERFORMANCE COMPARISON

We compare the above models based on some classification metrics: accuracy, precision, recall, and f1-score. We present a comparative analysis of different approaches we studied for predicting offender demographic from crime evidence and victim demographic data by plotting the results in radar charts. First, we compare the classical machine learning and deep learning algorithms we applied to our data. Then we take the best three models and compare them with the open-source models and our custom DNN model.

8 PERFORMANCE EVALUATION

This chapter discusses our experimental setup, data demography, and performance metrics. We show the performances of different models we used. Finally, we present an overall comparison of the results obtained.

8.1 Performance Metrics

We use four performance metrics for our study: accuracy, precision, recall, and f1-score. In this subsection, we give a general idea about these metrics.

8.1.1 Accuracy. Accuracy is the ratio of the number of correct predictions made by the model to all predictions made. We generally use this metric for balanced data. Accuracy is defined as follows [12]:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{3}$$

8.1.2 Precision. Precision is defined as the percentage of positive predictions that were accurately anticipated to all positive predictions. It attempts to respond to the following question: What proportion of positive predictions was actually valid? Precision is defined as follows [12]:

$$\text{Precision} = \frac{\text{Number of true positives}}{\text{Total number of positive predictions}} \tag{4}$$

8.1.3 Recall. Recall is the proportion of accurately anticipated positives to all actual positives. It attempts to answer the following question: What proportion of actual positives was identified correctly? Recall is defined as follows [12]:

$$\text{Recall} = \frac{\text{Number of true positives}}{\text{Total number of positives in the ground truth}} \tag{5}$$

8.1.4 F1-score. The F1-score combines the precision and recall of a classifier into a single metric by taking their harmonic mean. It is primarily used to compare the performance of two classifiers. F1-score is defined as follows [12]:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

8.2 Data Demography

Our datasets consist of mostly categorical features. These features are also nominal, which means there is no numerical relationship between different categories.

In the *Incident-Level Extract File* mentioned in subsection 4.2.1, all the categorical classes are presented as numbers for memory efficiency purposes. Detailed documentation is in the codebook provided with the data files. The NIBRS data received from the FBI contain blanks for missing data that sometimes might stand for a code with substantive meaning. In the extract files, all blanks are recoded to a negative integer.

We work with only single victim-single offender incidents in this study for simplicity. After feature selection, we use the following features to predict the offender demographic: *Crime Type, Type of Weapon Used, Incident Location, Type of Injury, Victim's Age Range, Victim's Gender, Victim's Race and Victim-Offender Relation.* As for offender demographic prediction, we try to predict the following features: *Offender's Age Range, Offender's Gender and Offender's Race.*

In the *NYPD Complaint Data* mentioned in subsection 4.2.2, each row is a complaint record. The missing values are represented with NULL. We use the following features to predict the offender demographic: *Offense Description, Incident Month, Incident Location, Victim's Age Range, Victim's Gender and Victim's Race.* As for offender demographic prediction, we try to predict the following features: *Suspect's Age Range, Suspect's Gender and Suspect's Race.*

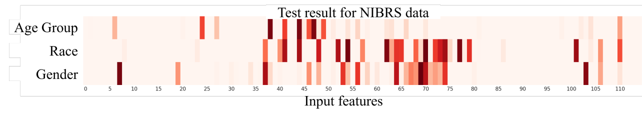

Figure 8: Heat map for NIBRS data

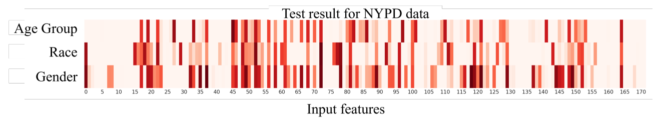

Figure 9: Heat map for NYPD complaint data

8.3 Experimental Setup

We use *pandas* for data preprocessing, which is a software library written for the Python programming language for data manipulation and analysis. We primarily use *Weka* for experimenting with machine learning models. It is a compilation of different machine learning methods used for data mining. Later, we shift to *Google Colab* for experimenting with deep learning models. We mainly used the *Keras* deep learning framework for designing our models. In the open-sourced model we explored in our study, they use *Theano* library. It is a Python library and optimizing compiler for handling and evaluating mathematical statements, particularly those with matrix-valued variables. In Theano, computations are defined using a syntax similar to NumPy and built for efficient execution on CPU and GPU architectures [11]. Nevertheless, the developers discontinued its update; thus, we reformed their code to some extent.

8.4 Chi-Square Test Results

We get the correlation information about input features and class labels from the Chi-Square test. We graphically present this information with heat maps. We can see from the heat maps in Figure 8 and 9 that the darker the color, the lesser the correlation between the two features. Depending on this property, we discard input features that have little or no correlation with the class labels. This results in lesser computation time with no significant change in performance.

8.5 Performance Comparison of Classical Machine Learning and Deep Learning Algorithms

In this subsection, we show the performance comparison of the classical machine learning and deep learning models we use in our study mentioned in subsection 5 and Subsubsection 6.1. We use the performance metrics mentioned in subsection 8.1 to evaluate the models. We present the results in radar charts for both datasets in Figure 10 and 11.

8.6 Performance Comparison of Open-Sourced and Custom Deep Learning Models

In this subsection, we show the performance comparison of the open-sourced deep learning models we use in our study mentioned

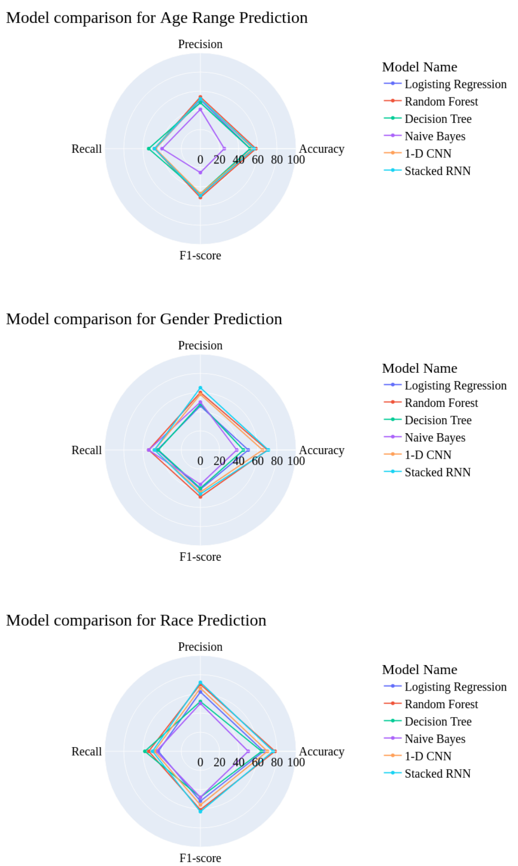

Figure 10: Performance comparison for NIBRS data

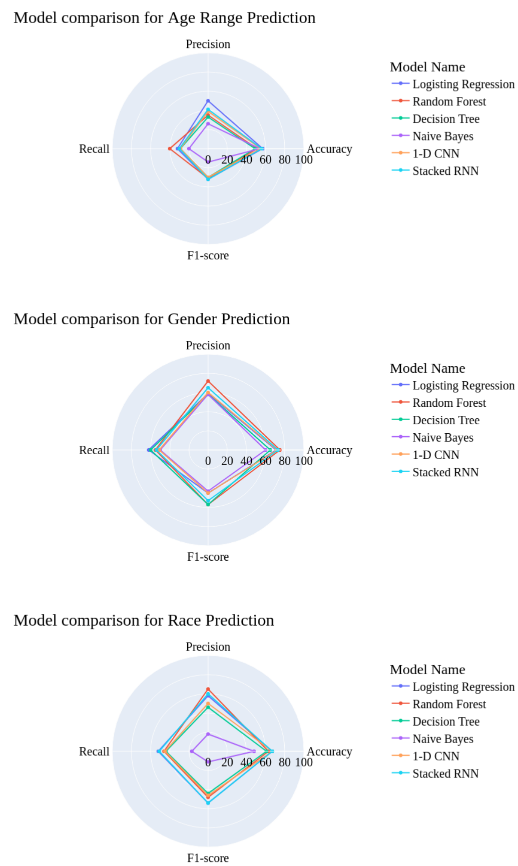

Figure 11: Performance comparison for NYPD complaint data

in Subsubsection 6.2 and our custom model mentioned in Subsubsection 6.3 with the best performed classical ML and DL models from subsection 8.5. We use the performance metrics mentioned in subsection 8.1 to evaluate the models. We present the results in radar charts for both datasets in Figure 12 and 13.

8.7 Comparative Analysis of the Results Obtained

From the radar charts shown in subsection 8.5, we can see that the overall better-performing models among the classical machine learning and deep learning models are: *Logistic Regression*, *Random Forest* and *Stacked Bidirectional LSTM*. We compare these three with the open-source models and our proposed DNN model. We can see from the radar charts in subsection 8.6 that the open-source models and our proposed DNN outperform the classical models in all the performance metrics. While predicting *Criminal Gender*, we can see that the open-source models achieve a high recall score. Nevertheless, the models become biased because the Criminal Gender attribute is biased. We try to tackle this issue by introducing *Class Weighted Loss* in our model. We lose a bit of recall score in this process, but the overall scores are balanced.

9 DISCUSSION

In this subsection, we discuss in detail how we answer our research questions and how our study compares with the existing studies. We also discuss the difficulties we faced in conducting our study.

9.1 Difficulties in Crime Data Collection and Lack of Availability of Good Datasets

It is a challenging task in and of itself to collect appropriate crime datasets. Since they are private and sensitive information, law enforcement organizations rarely publish the datasets. Even when published, a significant amount of the information that could have been useful for research is removed. Therefore, it is quite a challenge for individuals interested in conducting research in computer criminology to locate appropriate datasets for supporting their research ideas. Even though we use two extensive datasets in this study, they are only from the United States of America. We could not locate any Bangladeshi or international datasets that met our requirements. In addition, the datasets that we use have many missing values and a limited number of criminal characteristics. As a result, we can only work with the limited demographic features currently available.

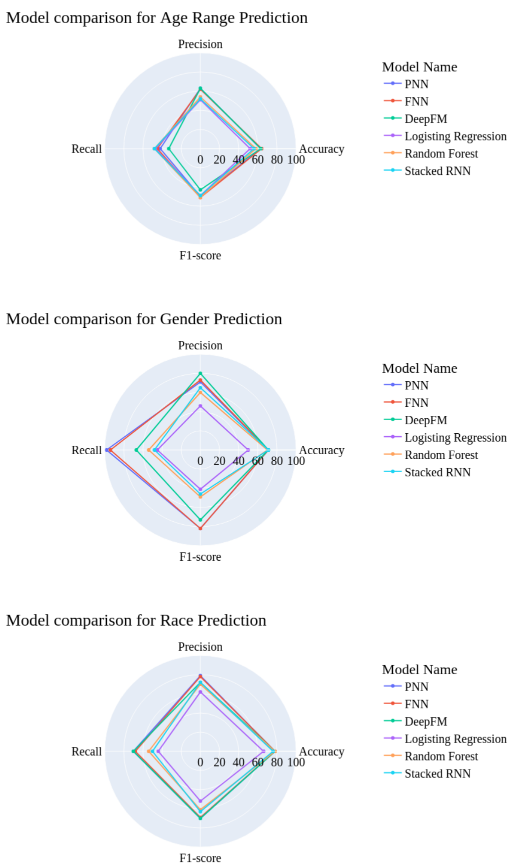

Figure 12: Performance comparison for NIBRS data

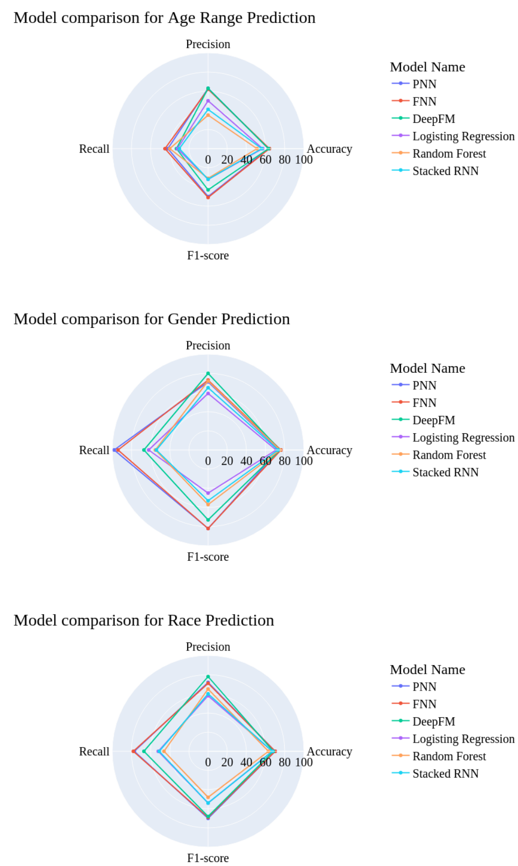

Figure 13: Performance comparison for NYPD complaint data

9.2 Prediction of Offender Demographic Features from Crime Evidence Data and Victim Demographic Features

We investigate a number of different machine learning algorithms and assess how well they function with our data. Logistic Regression, Decision Tree, Random Forest, and Naive Bayes are some of the algorithms that we use. In addition to this, we experiment with various deep learning algorithms to see if this helps improve the performance. In order to compare performances, we use both 1-D CNN and Stacked RNN. Due to the fact that all of our data were categorical, the models were unable to capture the feature interactions fully. Therefore, we search for open-source deep learning models that can function well with categorical data. In order to evaluate how well their models work, we fit our data into their models. We propose a custom DNN model and contrast our findings with theirs. In addition to this, we make use of weighted loss in an effort to eliminate model bias.

9.3 Comparison with Other Existing Studies

There are few studies regarding the machine learning approach in criminal profiling. The existing studies primarily work with specific crime types, serial killing, and criminal MO. We take a more generalized approach to offender demographic profiling. Also, we use two datasets on which this type of study has not been conducted yet. We also preprocess the NIBRS data, which is known to have a complex data structure. We try to give it a more straightforward representation to make it easier for machine learning models to work with this data.

10 FUTURE WORK

Our study has two fundamental limitations so far. One is the datasets we use. These datasets are prepared through an annual survey in the USA and have a lot of missing values. Depending on how one handles the missing values, results can vary immensely. Also, the data is entirely categorical and nominal. It is pretty hard for machine learning models to work with this type of data. Secondly, the acceptability of criminal profiling. Many investigators do not accept its credibility and have little faith in this procedure [30]. There is a chance that investigators will criticize it even if the procedure is backed by data analysis and machine learning techniques. We can apply different data imputation techniques in these datasets for our future work. It can be a research topic on its own, not an extension of this study. We can also shift our attention to improving the custom model we propose in this study to make better predictions. We can also use these datasets for demographic analysis, such as what kind of people are at higher risk of being assaulted, what kind of people are more likely to commit offenses and which area is riskier to live in than other areas.

11 CONCLUSION

Automated prediction of criminal demographic profiles in a generalized manner is little explored in the literature. Although advances in computer and information technologies have encouraged law enforcement agencies to compile extensive crime incident record databases with crime details, big data analysis, and deep learning techniques are not used much in this sector. To predict criminal demographic features from crime evidence data and victim demographic features, we examine a variety of machine learning and deep learning techniques. In addition to this, we propose a DeepFM-based DNN framework for the purpose of criminal demographic profiling. It is superior to both traditional machine learning algorithms and deep learning algorithms in terms of speed and accuracy. In exchange for a marginal decrease in performance, it is able to function normally despite the presence of unbalanced data. Hence, we can apply these methods to build a decision-aid tool to help investigators narrow down the list of suspects in unsolved crime cases. We intend to continue developing our DNN architecture in the hopes of enhancing its functionality. In addition to this, we intend to work on more crime datasets, focusing particularly on crime statistics from Bangladesh.

REFERENCES

[1] Lynn A. Addington. 2019. *NIBRS as the New Normal: What Fully Incident-Based Crime Data Mean for Researchers.* Springer International Publishing, Cham, 21–33. https://doi.org/10.1007/978-3-030-20779-3_2

[2] Yoshio Akiyama and James Nolan. 1999. Methods for Understanding and Analyzing NIBRS Data. *Journal of Quantitative Criminology* 15, 2 (01 Jun 1999), 225–238. https://doi.org/10.1023/A:1007531023247

[3] Abrar A. Almuhanna, Marwa M. Alrehili, Samah H. Alsubhi, and Liyakathunisa Syed. 2021. Prediction of Crime in Neighbourhoods of New York City using Spatial Data Analysis. In *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA).* IEEE, Riyadh, Saudi Arabia, 23–30. https://doi.org/10.1109/CAIDA51941.2021.9425120

[4] Khaled A. Althelaya, El-Sayed M. El-Alfy, and Salahadin Mohammed. 2018. Evaluation of bidirectional LSTM for short-and long-term stock market prediction. In *2018 9th International Conference on Information and Communication Systems (ICICS).* IEEE, Irbid, Jordan, 151–156. https://doi.org/10.1109/IACS.2018.8355458

[5] K.C. Baumgartner, S. Ferrari, and C.G. Salfati. 2005. Bayesian Network Modeling of Offender Behavior for Criminal Profiling. In *Proceedings of the 44th IEEE Conference on Decision and Control.* IEEE, Seville, Spain, 2702–2709. https://doi.org/10.1109/CDC.2005.1582571

[6] PATRICK E. COOK and DAYLE L. HINMAN. 1999. Criminal Profiling: Science and Art. *Journal of Contemporary Criminal Justice* 15, 3 (1999), 230–241. https://doi.org/10.1177/1043986299015003002 arXiv:https://doi.org/10.1177/1043986299015003002

[7] Christopher Devery. 2010. Criminal Profiling and Criminal Investigation. *Journal of Contemporary Criminal Justice* 26, 4 (2010), 393–409. https://doi.org/10.1177/1043986210377108 arXiv:https://doi.org/10.1177/1043986210377108

[8] Sampath Kumar Gajawada. 2019. Chi-square test for feature selection in machine learning. https://towardsdatascience.com/chi-square-test-for-feature-selection-in-machine-learning-206b1f0b8223#:~:text=A%20chi%2Dsquare%20test%20is,count%20O%20deviates%20each%20other.

[9] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. https://doi.org/10.48550/ARXIV.1703.04247

[10] Jeff Heaton. 2016. An empirical analysis of feature engineering for predictive modeling. In *SoutheastCon 2016.* IEEE, Norfolk, VA, USA, 1–6. https://doi.org/10.1109/SECON.2016.7506650

[11] Toan Hoang. 2019. Tableau and Python / An Introduction. [Online; accessed 2022-05-14].

[12] Mohammad Hossin and Md Nasir Sulaiman. 2015. A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process* 5, 2 (2015), 1.

[13] Uday Kamal, Shamir Ahmed, Tarik Reza Toha, Nafisa Islam, and A. B. M. Alim Al Islam. 2020. Intelligent Human Counting through Environmental Sensing in Closed Indoor Settings. *Mob. Netw. Appl.* 25, 2 (apr 2020), 474–490. https://doi.org/10.1007/s11036-019-01311-w

[14] Richard N. Kocsis. 2006. *What Is Criminal Profiling?* Humana Press, Totowa, NJ, 1–11. https://doi.org/10.1007/978-1-59745-109-3_1

[15] Jessie L. Krienert and Jeffrey A. Walsh. 2018. An Examination of Intimate Partner Sexual Violence: Comparing Marital and Nonmarital Incidents Employing NIBRS Data, 2008–2012. *Partner Abuse* 9, 1 (2018), 41–57. https://doi.org/10.1891/1946-6560.9.1.41 arXiv:https://connect.springerpub.com/content/sgrpa/9/1/41.full.pdf

[16] Yasmine Lamari, Bartol Freskura, Anass Abdessamad, Sarah Eichberg, and Simon de Bonviller. 2020. Predicting Spatial Crime Occurrences through an Efficient Ensemble-Learning Model. *ISPRS International Journal of Geo-Information* 9, 11 (2020). https://doi.org/10.3390/ijgi9110645

[17] Hairen Liao, Lingxiao Peng, Zhenchuan Liu, and Xuehua Shen. 2014. iPinYou global rtb bidding algorithm competition dataset. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising.* ACM, New York, United States, 1–6.

[18] Christina A. Mehranbod, Ariana N. Gobaud, and Christopher N. Morrison. 2022. Ridesharing and alcohol-related assaults in NYC: A spatial ecological case-crossover study. *Drug and Alcohol Dependence* 232 (2022), 109321. https://doi.org/10.1016/j.drugalcdep.2022.109321

[19] National Archive of Criminal Justice Data. 2021. National Incident-Based Reporting System, 2016: Extract Files.

[20] National Archive of Criminal Justice Data. 2022. National Incident-Based Reporting System Resource Guide. https://www.icpsr.umich.edu/web/pages/NACJD/NIBRS/

[21] Police Department (NYPD). 2022. Crime Stats - Historical - NYPD. https://www1.nyc.gov/site/nypd/stats/crime-statistics/historical.page

[22] Harshil Patel. 2021. What is feature engineering - importance, tools and techniques for machine learning. https://towardsdatascience.com/what-is-feature-engineering-importance-tools-and-techniques-for-machine-learning-2080b0269f10

[23] Police Department (NYPD). 2021. NYPD Complaint Data Historic | NYC Open Data. https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i

[24] Yanru Qu, Han Cai, Kan Ren, Weinan Zhang, Yong Yu, Ying Wen, and Jun Wang. 2016. Product-based Neural Networks for User Response Prediction. https://doi.org/10.48550/ARXIV.1611.00144

[25] Nachirat Rachburee and Wattana Punlumjeak. 2015. A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining. In *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE).* IEEE, Chiang Mai, Thailand, 420–424. https://doi.org/10.1109/ICITEED.2015.7408983

[26] Sunil Ray. 2017. Commonly Used Machine Learning Algorithms | Data Science. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

[27] Michael L. Rich. 2016. MACHINE LEARNING, AUTOMATED SUSPICION ALGORITHMS, AND THE FOURTH AMENDMENT. *University of Pennsylvania Law Review* 164, 4 (2016), 871–929. http://www.jstor.org/stable/24753528

[28] Shipra Saxena. 2020. Categorical Data Encoding Techniques | Analytics Vidhya. [Online; accessed 2022-05-14].

[29] Cedric Seger. 2018. An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. , 34 pages.

[30] Brent Snook, Joseph Eastwood, Paul Gendreau, Claire Goggin, and Richard M. Cullen. 2007. Taking Stock of Criminal Profiling: A Narrative Review and Meta-Analysis. *Criminal Justice and Behavior* 34, 4 (2007), 437–453. https://doi.org/10.1177/0093854806296925 arXiv:https://doi.org/10.1177/0093854806296925

[31] Scott J. South and Steven F. Messner. 2000. Crime and Demography: Multiple Linkages, Reciprocal Relations. *Annual Review of Sociology* 26 (2000), 83–106. http://www.jstor.org/stable/223438

[32] B. Venkatesh and J. Anuradha. 2019. A Review of Feature Selection and Its Methods. *Cybernetics and Information Technologies* 19, 1 (2019), 3–26. https://doi.org/10.2478/cait-2019-0001

[33] Wikipedia contributors. 2022. National Incident-Based Reporting System. https://en.wikipedia.org/wiki/National_Incident-Based_Reporting_System

[34] Weinan Zhang, Tianming Du, and Jun Wang. 2016. Deep Learning over Multi-field Categorical Data: A Case Study on User Response Prediction. https://doi.org/10.48550/ARXIV.1601.02376

[35] Yu-Jun Zheng, Wei-Guo Sheng, Xing-Ming Sun, and Sheng-Yong Chen. 2017. Airline Passenger Profiling Based on Fuzzy Deep Machine Learning. *IEEE Transactions on Neural Networks and Learning Systems* 28, 12 (2017), 2911–2923. https://doi.org/10.1109/TNNLS.2016.2609437