Assault Prediction

DATA MINING, PREDICTING ASSAULTS, AND COMMUNICATING RISK

Indiana Department of Correction
Bret Ellis, Research Scientist, & Aaron Garner, Executive Director Research and Technology,
Staff are our most valuable asset.

These individuals are committed to the betterment of society through rehabilitation and public safety. It is incumbent on us to provide the safest working environments possible.

Warning: Graphic
Introduction

Following a series of brutal staff assaults in late spring 2016, the Indiana Department of Correction began to search for additional ways to combat facility violence. Deputy Commissioner of Operations, James Basinger, initiated a scientific evaluation into the underlying causes of staff assault.

- Members of the IDOC Research & Planning data science team, Sarah Schelle & Bret Ellis, managed the analytics project.
- After discussing the problem with Operations staff, it became apparent that a tool designed to detect and communicate when offenders were at their most volatile would systematically help to combat staff assault.
Static versus Dynamic Specific to Your Outcome of Interest

Most risk assessments, due to the nature of when they are administered in the corrections onboarding experience, use a few static variables (ie: At what age was the offender’s first arrest? First tried drugs? Etc.) associated with one or two outcomes (Recidivism).

Fundamentally, we know that even the most dangerous individual does not commit an assault every day. Instead there are environmental or dynamic factors that act as triggers to make an act of violence more likely on a particular day.

By using data recorded in our offender information management systems we have access to dozens of risk variables being recorded and managed in near real time.

- By utilizing existing risk data conducted at onboarding, as well as real time system variables in our model, we get the best of both worlds.
Assault prediction and mitigation took place over 4 stages.

- **Stage 1**
  - Data reporting was delivered to Operations identifying offenders with certain non-scientifically validated variables (later validated). This included whether offenders had a history of staff assaults. This represents the first data intervention. July, 2016

- **Stage 2**
  - Model Development

- **Stage 3**
  - Model Refinement
  - Full Rollout (included training webinars and providing access to operations management) late Sept. 2016

- **Stage 4**
  - Policy changes as a result of knowledge gained from modeling process
A model consists of independent variables (predictors) and dependent variable(s) (outcomes)

\[ y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i, \]

Initial Stakeholder Meetings to assist in predictor variable development

Creation of modeling dataset (using 1 year failure periods) (2012-2015)

- Using Conduct Violations as outcome data
- And a variety of operational and risk variables as predictors (next page):

Over 100 variables from multiple information silos or mainframe extracts were tested in the course of the project.
Project Timeline
Stage 2 Model Development Process

Modeled using SAS Enterprise Miner and a variety of algorithms including:
- CHAID
- Auto Neural Network (machine learning and deep learning)
- Regression

Unexpected Results: Many of the traditional risk items were statistically non significant

- Current Age
- Security Classification
- Sentence Length
- Idle Status
- Housing Type
- Academic Classification
- Crime Category
- Most Serious Offense
- Security Threat Group Status
- Custody Classification
- Indiana Risk Assessment Score (IRAS)
- Duration to Release
- Mental Health Classification

*Only a sample of the variables are displayed.*
Project Timeline
Stage 2 Model Development Process

Unexpected Results

- Why were variables that we know from scientific literature to be important, not a significant predictor of staff assaults?
  - (Anyone?)
  - Because they are known and managed!
  - For example, we systematically deny opportunities for STG offenders to interact and engage in violent conduct.

This modeling process acts as an evaluation of your existing operational practices in relation to the outcome of interest.
After evaluating the models, a CHAID (decision tree) was chosen as best fit and for ease of communication

- A Two-fold validation model was used to evaluate model performance and select the CHAID model. Two fold validation involves splitting your dataset in half. Using the first half (training set) to develop the model, and using the second half (validation set) to test the fit of the model. This allows you to avoid overfitting.
Project Timeline
Stage 2 Model and Score

Using the associated training probabilities with each leaf of the model, a score is generated for each member of the current offender population on a weekly basis. (as we have non timely data scattered across multiple information silos, it is scored weekly. If a state had a single information warehouse, it would be easily possible to score in real time.)
SAS Visual Analytics Portal was used to create a risk communication tool.
Stage 3

- Custody leadership and selected staff are provided access to the SAS VA Assault Prediction Portal

Model Refinement - as focus groups continued with custody experts a variety of variables are suggested that are tested to determine if they improve the model.
Project Timeline
Stage 3 Model Development Process

- Current Age
- Total Conduct History with IDOC
- Sentence Length
- Idle Status
- Housing Type

- Mental Health Code Change
- Change to Idle *
- Recent Contraband Violations
- History of Assaults

- Indiana Risk Assessment Score
- Duration to Release
- Mental Health Classification
Modeling Results

Using our standard risk assessment, we are able to identify highest risk offenders who have 2.5 times the likelihood of committing an assault, as the largest population segment.

Using our new assaults model, we are able to identify highest risk offenders who have 41.5 times the likelihood of committing an assault, as the largest population segment.

We achieve far greater differentiation in assault likelihood using more dynamic variables in conjunction with our risk instrument and other static variables, than our risk instrument can begin to achieve on its own.
Modeling Results

Old Model of Highest Risk Offenders

New Model of Highest Risk Offenders

Incidence of assaultive behavior
Results of Model Implementation
Reduction Percentages

IDOC Averages

- #102 Staff Assault Violations
  - 01/15-06/16, 32 per month
  - Post Intervention period (07/16-12/16), 24 per month
  - 25% reduction

- Total Assault Conducts
  - 01/15-06/16, 320 per month
  - Post Intervention period (07/16-12/16), 265 per month
  - 17% reduction

Pilot Averages

- #102 Staff Assault Violations
  - 01/15-06/16, 6.55 per month
  - Post Intervention period (07/16-12/16), 4 per month
  - 39% reduction

- Total Assault Conducts
  - 01/15-06/16, 75 per month
  - Post Intervention period (07/16-12/16), 58 per month
  - 23% reduction

By raising our adoption rates and information access the rest of IDOC should be able to achieve reductions on par with the pilot facilities.
IDOC Total Violence

- Pilot Implementation
- Full Rollout
Stage 3 Model Refinement & Lessons

Refinement
- Due to the relatively low incidence rate of staff assault (about 1 in 100 offenders commits a staff assault in any particular year), we actually ended up moving to predicting any violent act (inmate on inmate, inmate on staff, etc.). The greater incidence rate allows for greater segmentation in the prediction model.

Lessons Learned
- Using outcomes as predictors
  - We found that it is very easy, dependent upon how you do your data joins, or integrate data from other systems, to accidentally catch data that occurred after an assault (For Example: recording a job loss that occurred as a result of an assault on staff, as a predating the assault on staff. It is very important to make sure that you are aware of any lag in data entry.
- Data Timeliness
  - Do not model data variables that you will not have access to score in a timely manner. We had several variables that were in systems that we could not score in real/near real time. For example: visitor data could not be retrieved in a timely manner.
Stage 4
Implementation of Policy Changes Based off Modeling Results

Expected April- July 2017
- Utilizations of High Risk Offender Flag within the IDOC Case Management system
- Programmatic and Mental Health Step Down Units
- Frequent Unit Briefings with mental health staff on medication compliance and other MH issues
- Expanded education of staff (including contract staff) on mental health issues and management techniques
- Additional staff assault training with video review of recent assaults

Ongoing
- Contraband Reduction
  - Audit of visitation rooms for contraband movement reduction (Spoke design)
  - Substitution of additional Intelligence and Investigation Staff for vacant custody positions
  - Implementation of Fencing with Modern Camera Systems

We expect additional violence reduction as a result of these policy changes and increased adoption of the modeling results.
Likely Future Projects

Contraband Prediction
- Using more extensive visitation, strategic threat group, work crew, and housing unit data

Parole Violation for a New Crime
- Using dynamic parole data, and case notes text mining

Likely victims of staff assault
- Making extensive use of training data
For More Information

Aaron Garner
Executive Director Research and Technology
agarner@idoc.in.gov

Bret Ellis
Research Scientist
bellis@idoc.in.gov

Sarah Schelle
Research Scientist
sschelle@idoc.in.gov