Predictive Analytics and Crime Data

Business and Data Understanding

Predictive analytics is widely used in business applications to strengthen various departments for the ultimate goal of profit – time saving resources in finance, economic patterns and trends, marketing focus groups, accounting automation, etc. One area that has grown a lot of attention is crime, where predictive tools have gathered a lot of attention in predicting criminal activity based on various traits. Not only does this assist in tackling actions on the street-level, such as burglary, but it can also aid in “behind the scenes” crimes such as fraud. The analytical tools gives the power to split up time and resources for very specific instances of crime, which allow police and other legal officers to know which areas to focus on. For instance, a beat cop can identify which streets or areas are more likely to involve crime, and spend their time focused on that, or an auditor can look up different transactions for fraudulent activity.

Data Preparation and Usage

The specific use case from the text involves the Philadelphia Adult Probation and Parole Department. This department hired a statistics and criminology professor from the University of Pennsylvania to form the models that judges would use in making courtroom decisions based on the evidence and data. To prepare the data, this department had to figure out data that already existed which they could use to their advantage. When gathering the information, certain districts realized that they had collected a lot more data and stored in electronically than they realized. This included activity from different criminals an their history in the courtroom, records from Philadelphia prisons, and other police write-ups. A “risk-prediction” tool was created which mined data from several different databased. The researchers then started to test various prediction variables to find the highest accuracy in building the model, which is described in more detail in the next section. They found that very strong predictors included a variable that quantified the risk level for people to reoffend, the number of prior times they were in jail, the zip code of the criminal, and how long ago (in years) the crime was committed.

Model Searching and Scoring

After collecting this information and testing the different predictors, the researchers started a forecasting method. They decided to set a time when the forecasting would start, which could be any point in the “life cycle” of the criminal activity – i.e. when bail is set, when the crime is committed, etc., and another variable for when that ends. They used random forest modeling in this process with hundreds of decisions trees that each had a certain level of risk for someone who is new to committing a crime. The researchers were responsible for gathering the units of prediction, figuring out a time horizon, and then choose which specific outcome the forecasting tool would predict. As much as they were involved in this creation, it was up to those in the field to choose how the resources would be allotted to focus on these categories. They studied the number of staff available, various levels of caseloads and work assigned, and figured out a percentage breakdown of the work – 15% dedicated to probationers only in the high risk area, 25 to 30% in the medium risk, and 55-60% in the low risk category.

Model Application and Testing  
After applying the random forest model in the Philadelphia court system, it was recorded and used for a full year and ended up with a 66% accuracy in different categories (low, medium, high) risk offenses. The researchers created separated descriptions for all three levels of risk and accuracies for each one, along with predictions in each sub category. They found, for instance, that criminals on probation who had a high-risk level were over ten times more likely to commit a crime within a two-year increment of the forecast period, compared to low or medium levels. This was mapped out over a ten year period of data to identify the long term effects of those on probation.

Summary and Conclusion

The goal of this type of research was to aid police officers and the court system in identifying trends in criminal activity to help better manage time and resources. An emphasis was placed on collaboration, where the professor and research team worked with the court and police department in repeating the process many times to make the model as accurate as possible. As they found answers in the data, one of the most surprising elements was the high percentage of high-risk offenders in repeat activity, which had increased from the limited prediction that had been done previously. The legal workers were not equipped to handle such a drastic increase in high-risk activity, as the model had predicted. This led to various methods re-organizing caseloads so that any high-risk criminals or probationers could be handled accordingly. The reality is, there is only a limited amount of money to be applied, no matter how committed the system is in stopping crime. The researchers therefore implemented this fact within the tools they were using, giving a level of creativity to the predictions that allowed those in the criminal justice system to start looking at the data more seriously. This only supports the point that predictive analytics tools will continue to find involvement in the court system, and with one type of model (random forests) applied and conducted, there are many more experiments waiting to be tested to strengthen the accuracy of the predictions.

Sources

<https://www.ncjrs.gov/pdffiles1/nij/240696.pdf>

<https://courts.phila.gov/pdf/site/appd.pdf>

<https://www.nytimes.com/2020/02/06/technology/predictive-algorithms-crime.html>