WILLIAMSON, DOUGLAS

From: WILLIAMSON, DOUGLAS
Sent: Wednesday, January 20, 2016 11:01 AM
To: SINGLETON, BENJAMIN; CASTRO, CLAUDIA; LEVINE, EVAN; CHOHLASWOOD, ALEXANDER; NEUSTETER, REBECCA; JOY, MICHAEL; FREER, JOSEPH
Cc: BELLO, FRANK
Subject: RE: Predictive Policing Evaluation
Attachments: Re: Invitation to Participate in an Evaluation of Predictive Forecasting of Crime Demonstration Project--Important Dated Material that is due on October 1, 2015

Hi Ben.
Thanks for giving this so much thought. This is very helpful.
I am attaching a message OMAP sent to contracts a while back in response to some questions from the vendors.
Based on our original response and your message below I have a couple of comments...

- NYPD informed the vendors that we would be providing a grid to serve as the base for their predictions (This is one reason I was concerned about Keystats proposal yesterday to use census Block groups).
- Related to the number of boxes, in earlier correspondence NYPD requested that they predict 1% of the land area for EACH borough. We used one percent based on previous studies and this was deemed an operationally, reasonable amount of ground to cover. Using a 300ft grid for the ENTIRE city yields 98,208 boxes... 98028 * 0.01 = 982 boxes. Multiplying this by 3 (one for each tour) would be 2946 per day. We also requested predictions for six crime types plus one set of predictions for a weighted composite of the six.
- In addition to your metrics, I would also propose including the Prediction Accuracy Index (PAI) for comparisons which is essentially the hit rate divided by the prediction area. This has been one of the standard metrics in the crime mapping literature. http://www.palgrave-journals.com/sj/journal/v21/n1/full/8350066a.html

Looking forward to continuing the conversation...
Doug

From: SINGLETON, BENJAMIN
Sent: Wednesday, January 20, 2016 10:14 AM
To: CASTRO, CLAUDIA; LEVINE, EVAN; CHOHLASWOOD, ALEXANDER; NEUSTETER, REBECCA; WILLIAMSON, DOUGLAS; JOY, MICHAEL; FREER, JOSEPH
Cc: BELLO, FRANK
Subject: RE: Predictive Policing Evaluation

Hi all,

I just wanted to continue the conversation on evaluation metrics for the vendors that will continue through the demonstration. Here is what I propose although I'm not attached to any of these suggestions – I just wanted to get a conversation going. It might be best to gather again as a group to discuss. Also note – these are evaluation metrics to determine how well the predictive model performed. It does not reflect many of the other important considerations such as security, transparency into the algorithm, time to deployment, etc.

- 45 day evaluation (30 days feels too short, 60-90 days too long)
- All vendors predict using same raster grid (300' x 300' boxes or should they be bigger? PredPol defaults to 500'x500' by comparison)
    - Assign each raster grid box an Id which will be used for evaluation purposes.

- Must predict 4 boxes per precinct per platoon (77 precincts * 3 platoons = 924 box predictions per 24 hr period)
    - Alternatively, 1 box per sector per tour (about 787 sectors * 3 platoons = 2,361 box predictions per 24 hr period)
    - Note – raster grid can't overlap precinct or sector boundaries if we want the vendors to predict within a precinct or sector. (if our raster grid box overlaps 23/25 pcts, then that will cause issues, right?)
- The vendors would receive crime complaint data for the previous day (0000-2400) at approximately 0230 when the Crime Data Warehouse has all the data for the previous day. (need to set up job to handle this automatically)
- Must receive predictions by 8am (start of second tour) for the next 24 hours (e.g. predict for 2nd platoon, 3rd platoon and 1st platoon of following day) in CSV file format.
    - CSV fields: Id, CrimeOccured, CrimeProba, CrimeCount
        - Id – reference to raster grid box (makes it easy to look up and do evaluation)
        - CrimeOccurred - Did a crime happen in the box or not? (accepted values: 1 or 0)
        - CrimeProba - What is the probability that a crime happened in the box? (accepted values: float between 0-1)
        - CrimeCount - How many crimes happened in the box? (accepted values: >= 0)
- Model evaluation metrics:
    - CrimeOccurred (binary classification)
        - Precision – true positives over the number of true positives plus the number of false positives
        - Recall – true positives over the number of true positives plus the number of false negatives
        - Confusion matrix – aka contingency table – gives transparency to how well the algorithm performed
    - CrimeProba (binary classification)
        - Log loss – measurement of accuracy that incorporates the idea of probabilistic confidence
        - ROC AUC – area under receiver operating characteristic, statistic used for model comparison
    - CrimeCount (regression)
        - Root mean squared error (RMSE) – measure of the differences between values predicted by a model and the values actually observed

Best,

Ben

Benjamin Singleton
Administrative Staff Analyst
Office of the Deputy Commissioner Operations
New York Police Department
One Police Plaza, Room 1112
New York, NY 10038
Office: (646) 610-8889
Job Cell: (646) 300-2929 | Personal: (646) 241-7184

From: CASTRO, CLAUDIA
Sent: Wednesday, January 20, 2016 8:47 AM
To: LEVINE, EVAN; CHOHLASWOOD, ALEXANDER; NEUSTETER, REBECCA; WILLIAMSON, DOUGLAS; SINGLETON, BENJAMIN; JOY, MICHAEL; FREER, JOSEPH
Cc: BELLO, FRANK
Subject: RE: Predictive Policing Evaluation

Good Morning,,
As discussed, please see attached evaluation sheet.