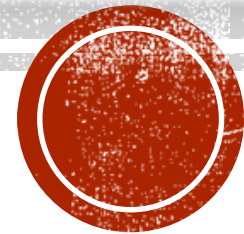


CHICAGO CRIME FORECAST



It's really not that bad

Varun Gupta – MSiA423



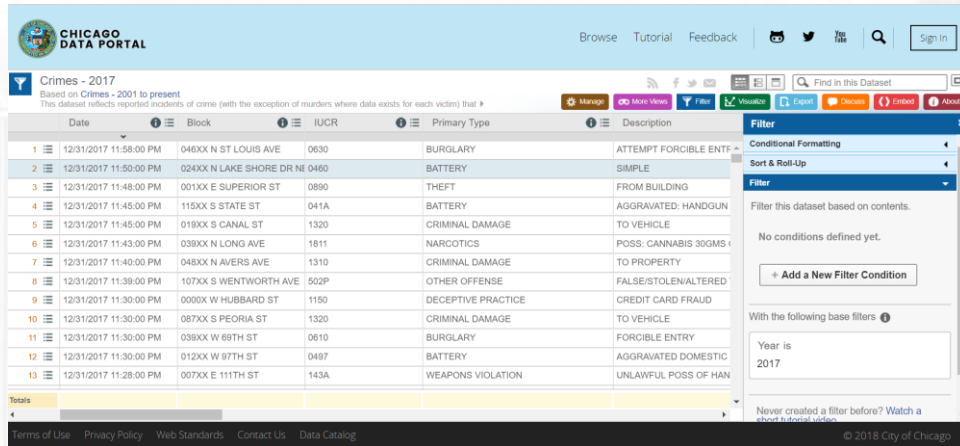
MOTIVATION

1. Provide means to compare crime rates in different neighborhoods of Chicago, especially for visitors who want to look up zip codes.
2. Explore the curious relationship, even studied by the Chicago Tribune, between crime and the weather.
3. Opportunity to explore Chicago's public data portal.
4. **Clear MSiA423**

<http://ec2-18-188-134-206.us-east-2.compute.amazonaws.com:5000/>



DATA STUFF

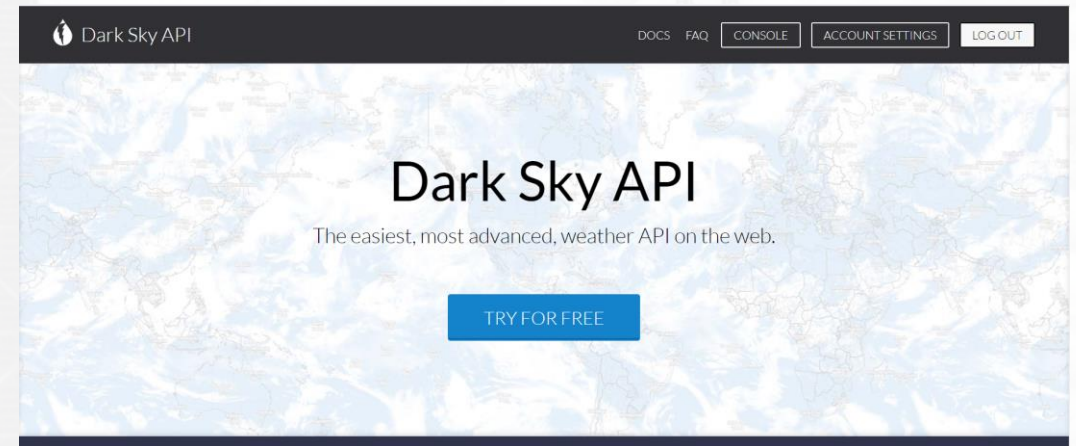


The screenshot shows the City of Chicago Data Portal interface. At the top, there's a header with the Chicago Data Portal logo and navigation links: Browse, Tutorial, Feedback, and social media icons. Below the header, a section titled 'Crimes - 2017' indicates the data is based on crimes from 2001 to the present. A table of crime data is displayed with columns for Date, Block, IUCR, Primary Type, and Description. The table lists 13 rows of data, including details like date, time, location (Block and IUCR), and crime type (Primary Type and Description). A 'Filter' sidebar is visible on the right, showing options for Conditional Formatting, Sort & Roll-Up, and a Filter section with a 'Filter' button. The bottom of the page includes links for Terms of Use, Privacy Policy, Web Standards, Contact Us, and Data Catalog, along with a copyright notice for 2018 City of Chicago.

	Date	Block	IUCR	Primary Type	Description
1	12/31/2017 11:58:00 PM	046XX N ST LOUIS AVE	0630	BURGLARY	ATTEMPT FORCIBLE ENTR
2	12/31/2017 11:50:00 PM	024XX N LAKE SHORE DR NE	0460	BATTERY	SIMPLE
3	12/31/2017 11:48:00 PM	001XX E SUPERIOR ST	0890	THEFT	FROM BUILDING
4	12/31/2017 11:45:00 PM	115XX S STATE ST	041A	BATTERY	AGGRAVATED: HANDGUN
5	12/31/2017 11:45:00 PM	019XX S CANAL ST	1320	CRIMINAL DAMAGE	TO VEHICLE
6	12/31/2017 11:43:00 PM	039XX N LONG AVE	1811	NARCOTICS	POSS: CANNABIS 30GMS
7	12/31/2017 11:40:00 PM	046XX N AVERS AVE	1310	CRIMINAL DAMAGE	TO PROPERTY
8	12/31/2017 11:39:00 PM	107XX S WENTWORTH AVE	502P	OTHER OFFENSE	FALSE/STOLEN/ALTERED
9	12/31/2017 11:30:00 PM	000XX W HUBBARD ST	1150	DECEPTIVE PRACTICE	CREDIT CARD FRAUD
10	12/31/2017 11:30:00 PM	087XX S PEORIA ST	1320	CRIMINAL DAMAGE	TO VEHICLE
11	12/31/2017 11:30:00 PM	039XX W 69TH ST	0610	BURGLARY	FORCIBLE ENTRY
12	12/31/2017 11:30:00 PM	012XX W 97TH ST	0497	BATTERY	AGGRAVATED DOMESTIC
13	12/31/2017 11:28:00 PM	007XX E 111TH ST	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HAN

City of Chicago Data Portal

- Historical Crime Data with Ward Information dating from Jan 2011- Feb2018. Data included date, ward, crime type, category, etc.
- Official Ward to Zipcode mapping data
- Historical Weather data for Chicago on each day across these years, courtesy Dark Sky API.
- Forecast data from Dark Sky, pulled multiple times a day using Cron to store it in RDS.

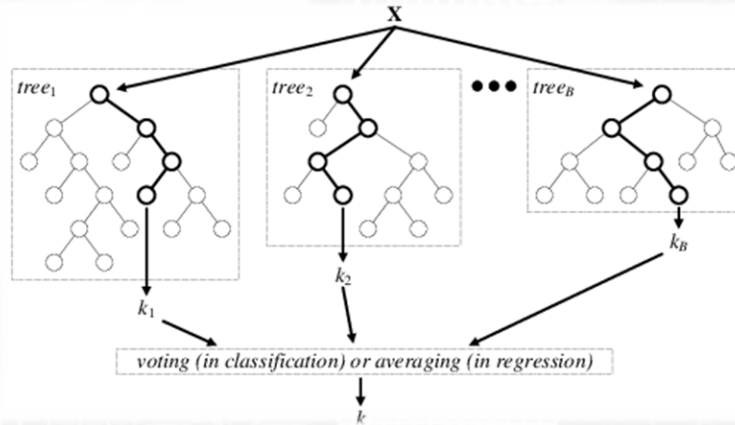


Dark Sky API



MODEL/ALGORITHM

- Data aggregated by day and zip-code, joined with weather data. Response variable for regression was total crime count.



- Finally used Random Forest model (Looking to upgrade this).
 - Why? Trees handle categorical variables with many levels very well.
 - Ensembles strongly improve variance without much increase in bias.
 - Success Criterion - Tuned for cross-validated SSE in crime count prediction.
 - Tuned parameters – 950 trees, max depth of 10



INSIGHTS AND RESULTS

- Chicago Data Portal -> Amazing Initiative
- ***Apparent*** temperature more significant than real temperature -> Used apparent temp. for model and predictions.
- Zipcode, temperature and precipitation intensity were all very significant predictors.

Other Insights:

- Insight – APIs can go sideways and leave you depressed.
- Scikit-learn doesn't do trees right.
- *Insight* – I can myself make a working web application that I can see through my browser using an actual URL: Mind=blown



QUESTIONS?

THANK YOU!

- Contact:
 - Email: varungupta2018@u.northwestern.edu
 - Phone: +1-312-964-0262
 - Github: varungupta93



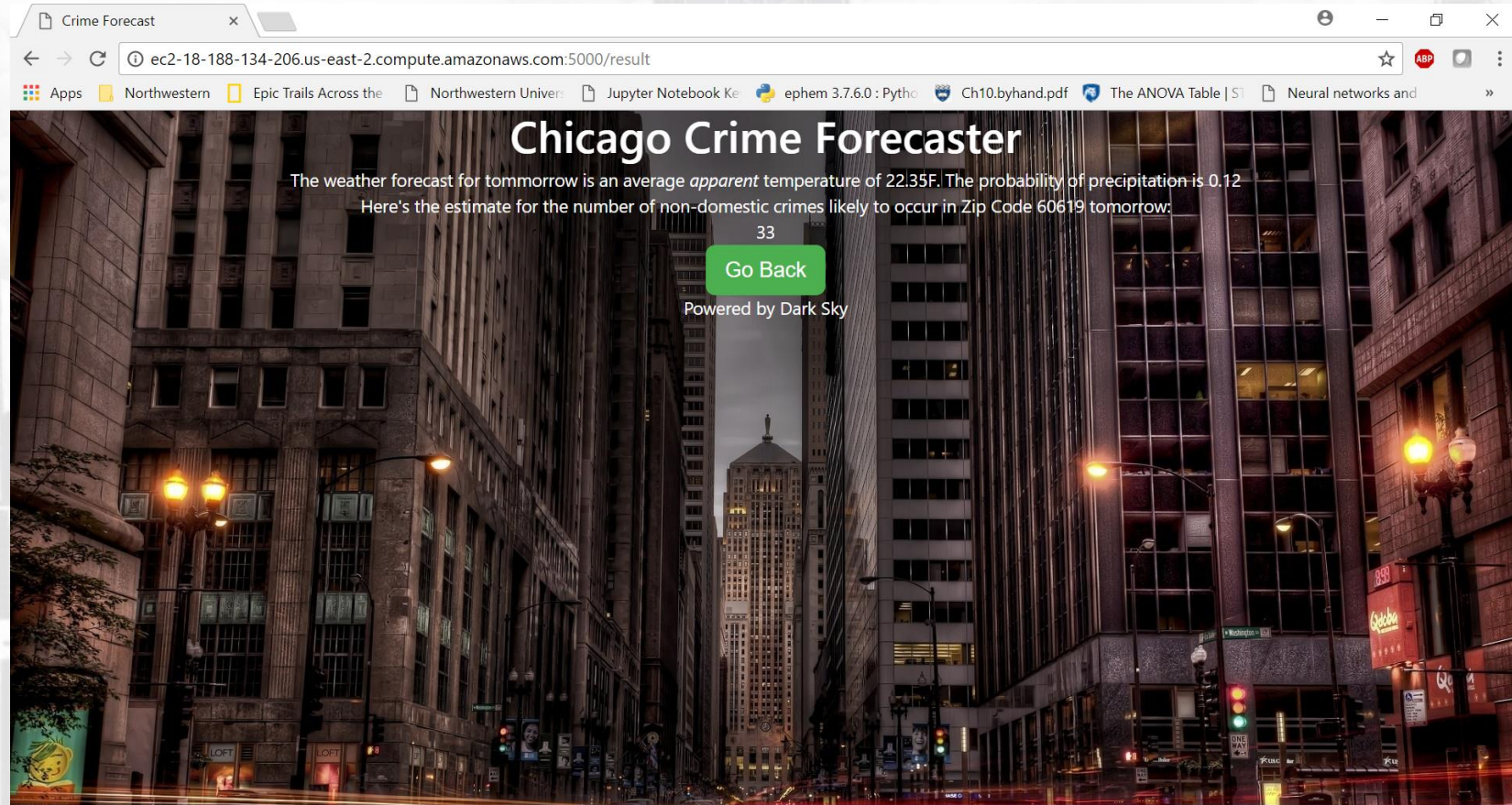
SCREENSHOTS



Minimalist Home Page



SCREENSHOTS



Results page (Maybe don't go to this Zipcode)

