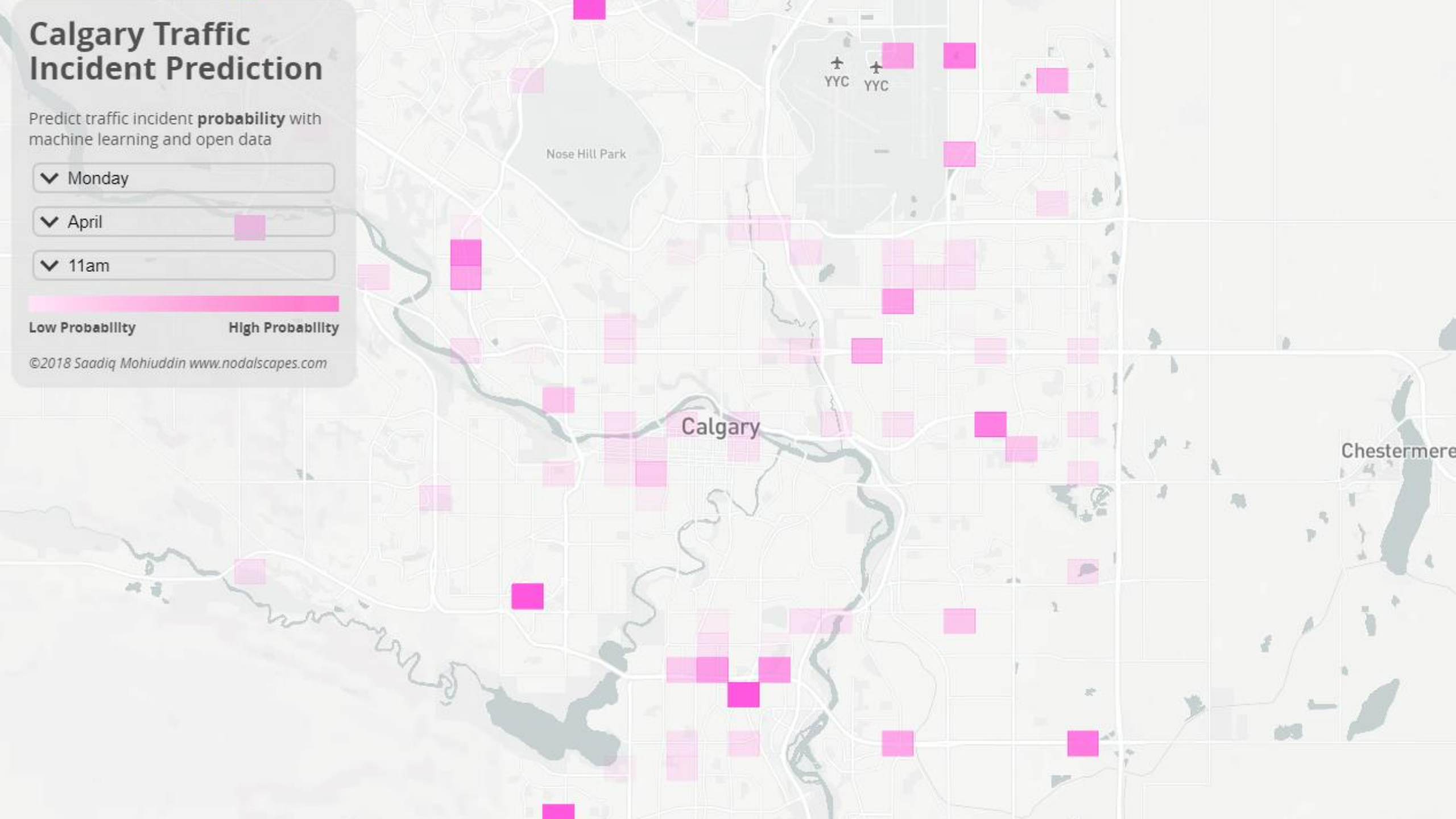
Predicting and Visualizing Traffic Collisions in Calgary with Machine Learning

Saadiq Mohiuddin

@saadiqmohiuddin

Github: https://github.com/smohiudd


Calgary Traffic Incident Prediction
Predict traffic incident **probability** with machine learning and open data
Monday
April
11am
Low Probability
High Probability
©2018 Saadiq Mohiuddin www.nodalscapes.com
YYC
YYC
Nose Hill Park
Calgary
Chestermere

Predicting Traffic Accidents Through Heterogeneous Urban Data: A Case Study

Zhuoning Yuan
University of Iowa
zhuoning-yuan@uiowa.edu

Xun Zhou
University of Iowa
xun-zhou@uiowa.edu

Tianbao Yang
University of Iowa
tianbao-yang@uiowa.edu

James Tamerius
University of Iowa
james-tamerius@uiowa.edu

Ricardo Mantilla
University of Iowa
ricardo-mantilla@uiowa.edu

ABSTRACT

With the urbanization process around the globe, traffic accidents have undergone a rapid growth in recent decades, causing significant life and property losses. Predicting traffic accidents is a crucial problem to improving transportation and public safety as well as safe routing. However, the problem is also challenging due to the imbalanced classes, spatial heterogeneity, and the non-linear relationship between dependent and independent variables. Most previous research on traffic accident prediction conducted by domain researchers simply applied classical prediction models on limited data without addressing the above challenges properly, thus leading to unsatisfactory performance. This paper, through a case study, presents our explorations on effective techniques to address the above challenges for better prediction results. Specifically, we formulate the problem as a binary classification problem. For each road segment in each hour, we predict whether an accident will occur. Big data including all the motor vehicle crashes from 2006 to 2013 in the state of Iowa, detailed road network, and various weather attributes at 1-hour granularity have been collected and map-matched. We evaluate four classification models, i.e., Support Vector

KEYWORDS

traffic accident prediction; big data; eigen-analysis; spatial heterogeneity

ACM Reference format:
Zhuoning Yuan, Xun Zhou, Tianbao Yang, James Tamerius, and Ricardo Mantilla. 2017. Predicting Traffic Accidents Through Heterogeneous Urban Data: A Case Study. In *Proceedings of 6th International Workshop on Urban Computing, Halifax, Nova Scotia, Canada, August 2017 (UrbComp-2017),* 9 pages.
https://doi.org/10.475/123_4

1 INTRODUCTION

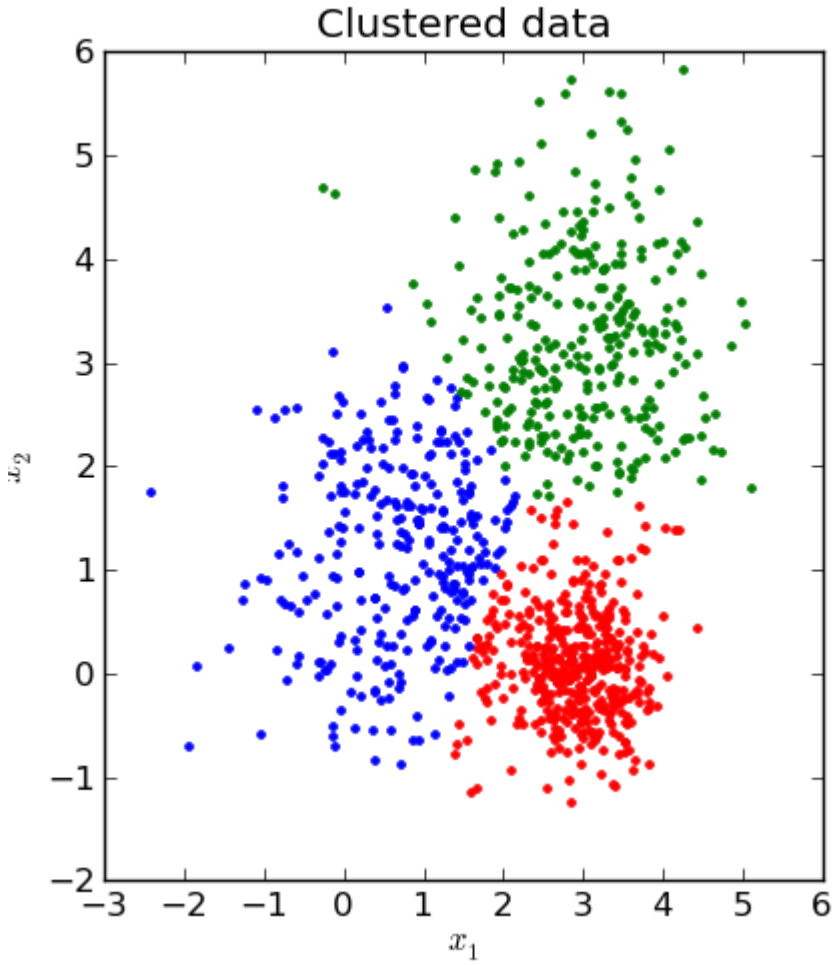
Traffic accidents have been a significant issue to public safety. The total traffic crash deaths reached 1.25 million in 2013 globally [25]. The ability to predict future accidents (e.g., where, when, or how) is thus very useful not only to public safety stakeholders (e.g., police) but also transportation administrators and individual travelers. A potential application of such technique would be real-time safe route recommendation for drivers. With the rapid development of data collection techniques and the availability of big urban datasets in recent years, predicting traffic accidents has become more realistic

Machine Learning

- Field of study that give computers the ability to learn without being explicitly programmed

- Unsupervised Learning:
  - Unlabeled data fed into algorithm which determines relationships/patterns in the data

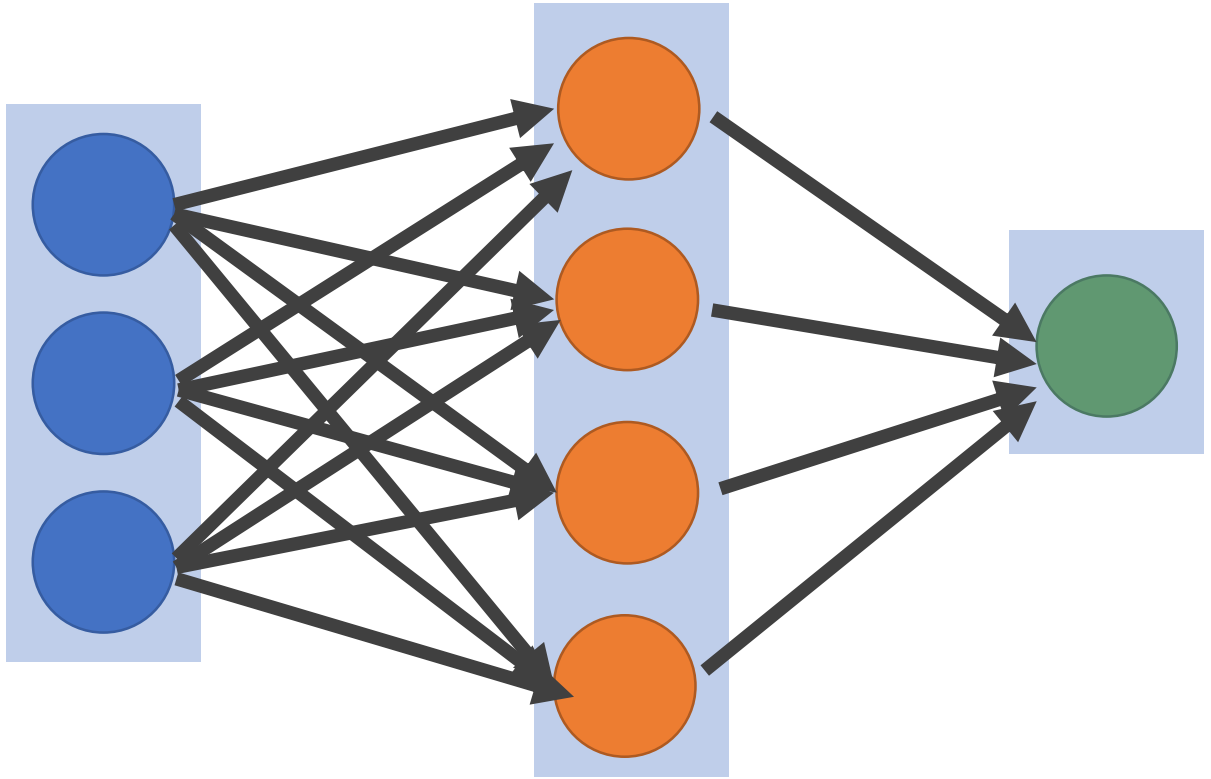
- Supervised Learning:
  - Given a dataset and known outputs (labelled data), we find a relationship between inputs and outputs



Artificial Neural Networks (ANN)/Deep Learning

- ANN inspired by biological neural networks
- Collection of connected “nodes” (loosely modeled on neurons)

Input Image

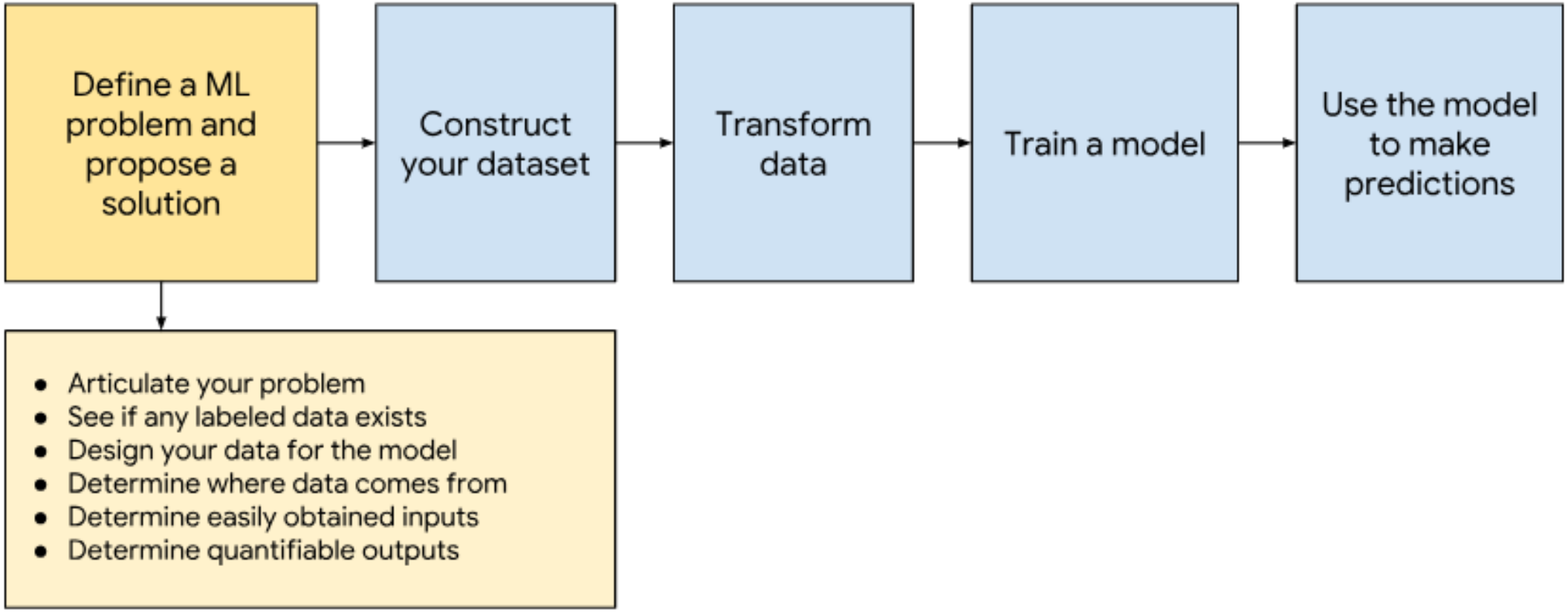


Cat
Or
Not a Cat

Getting Started

Make a Hypothesis

Test Hypothesis



Source: developers.google.com/machine-learning


jupyter

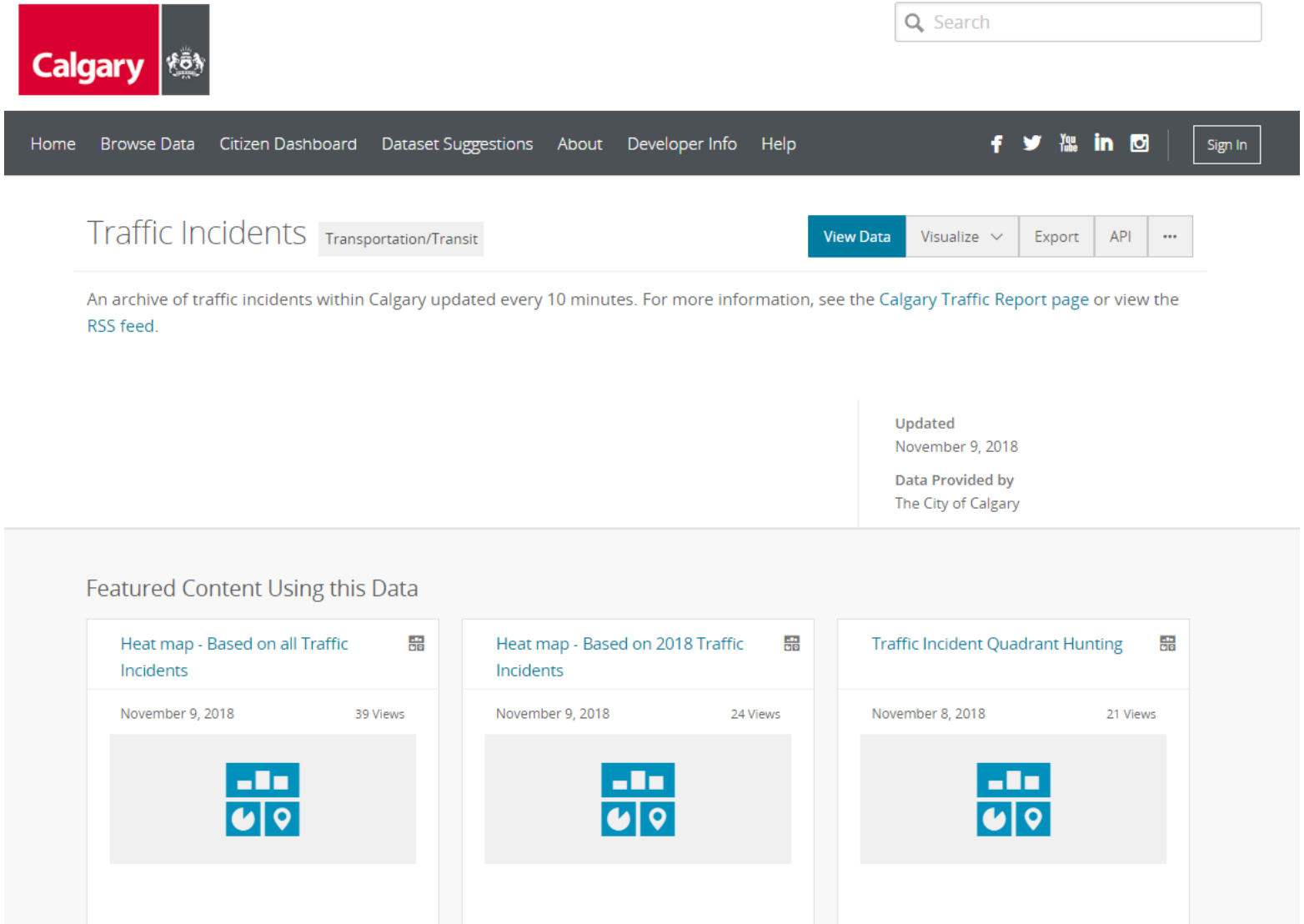

TensorFlow.js


K
Keras


React


mapbox

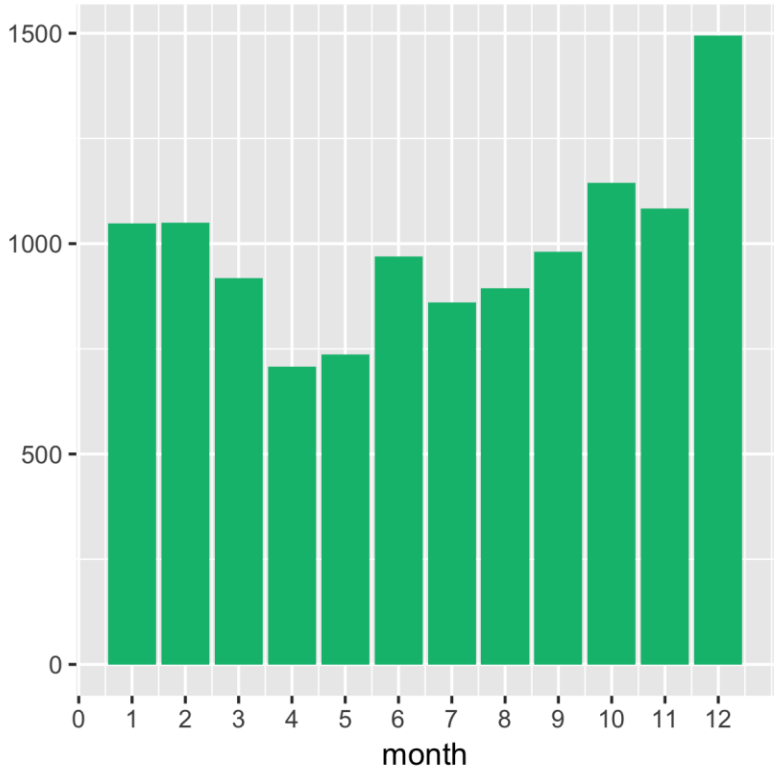
Data Source

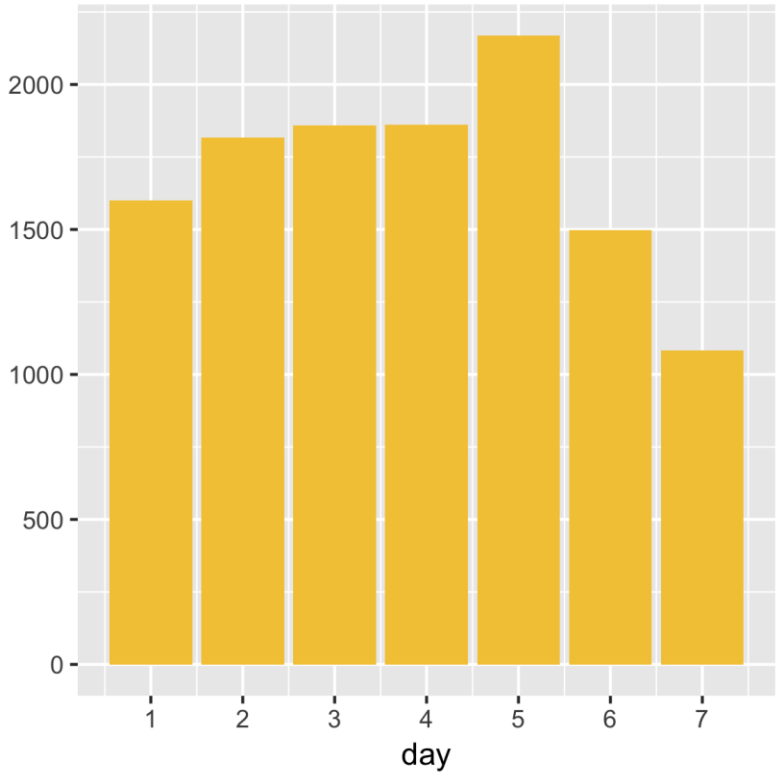

Calgary
Search
Home
Browse Data
Citizen Dashboard
Dataset Suggestions
About
Developer Info
Help
Sign In
Traffic Incidents
Transportation/Transit
View Data
Visualize
Export
API
An archive of traffic incidents within Calgary updated every 10 minutes. For more information, see the Calgary Traffic Report page or view the RSS feed.
Updated
November 9, 2018
Data Provided by
The City of Calgary
Featured Content Using this Data
Heat map - Based on all Traffic Incidents
November 9, 2018
39 Views
Heat map - Based on 2018 Traffic Incidents
November 9, 2018
24 Views
Traffic Incident Quadrant Hunting
November 8, 2018
21 Views

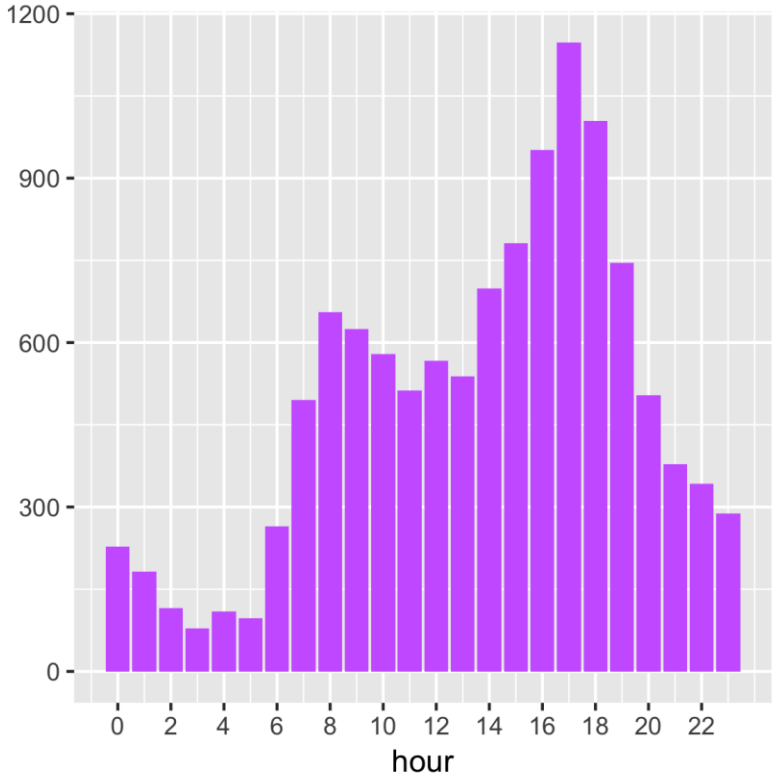
Data Exploration

```
> head(incident_types,20)
# A tibble: 20 x 1
   description
   <chr>
 1 Traffic signals are flashing red. Crews have been dispatched.
 2 Two vehicle incident.
 3 Multi-vehicle incident. blocking the right lane
 4 Two vehicle incident. blocking the left lane
 5 Single vehicle incident.
 6 Multi-vehicle incident. Blocking the middle lanes.
 7 There is an incident involving a pedestrian.
 8 Single vehicle incident. The road is closed btw Aero Gate and 19 St
 9 Two vehicle incident. blocking the right lane
10 Two vehicle incident. blocking the right lane on the ramp to NB
11 Single vehicle incident. blocking the left lane
12 Multi-vehicle incident.
13 Two vehicle incident. Moved to the right lane.
14 Stalled vehicle. Blocking the left lane.
15 Test incident
16 Traffic signals are flashing red.  Crews have been dispatched
17 Two vehicle incident. Blocking the middle lane
18 The traffic signals are blank. Crews have been dispatched.
19 the road is closed between 39 Avenue and 32 Avenue NE. Due to a police incident.
20 Two vehicle incident. Blocking the right lane
```

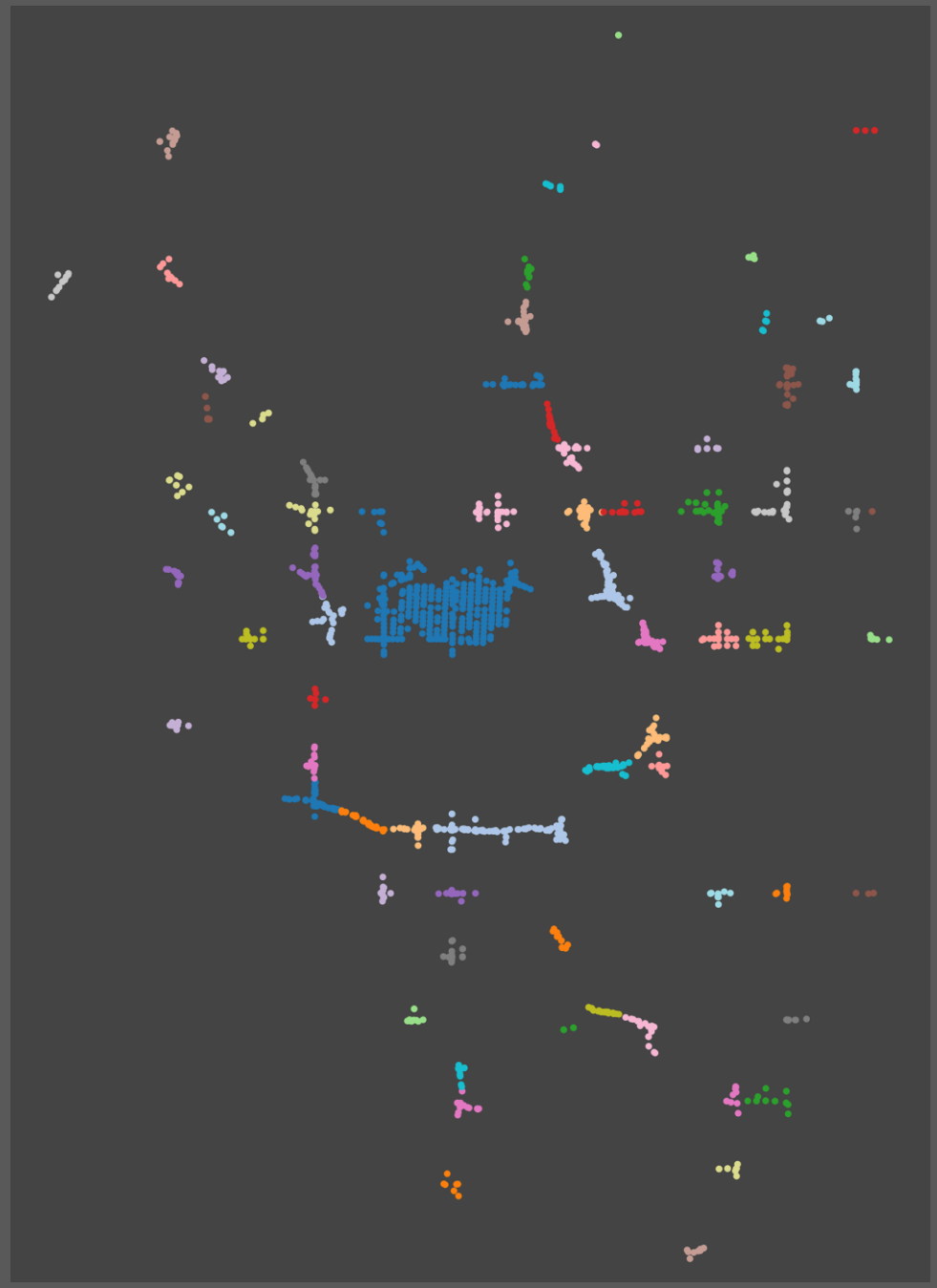
Data Exploration








Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Jan
Feb
Mar
Apr
May
June
July
Aug
Sept
Oct
Nov
Dec



Data Bias/Missing Data Sets

- How is the data collected?
- Is there missing data? What data is missing?
- How is the model being used?
- Can the code and data be audited? Is it open source?

Feature Engineering

- Most of your time with ML will involve feature engineering

- *“Feature engineering is the process of transforming **raw data** into **features** that better represent **the underlying problem** to **the predictive models**”*

  (Jason Brownlee, machinelearningmastery.com)

Feature Engineering

2018-11-13T19:03:13+00:00

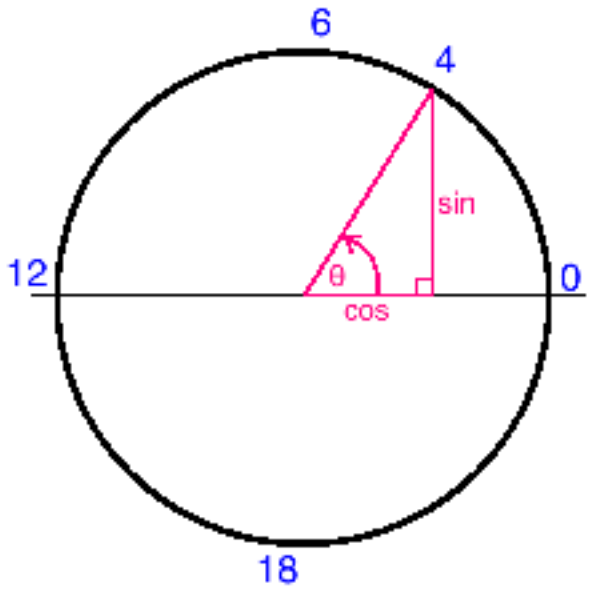
- Hour
- Minute
- Day
- Day of Week
- Year
- Holiday
- Time to Sunrise
- Temperature
- Precipitation
- Wind Speed

Lat/Lon Features

Out[5]:

date	start_dt	geohash
2017-11-15	2017-11-15 17:26:10	c3nfhj
2018-09-01	2018-09-01 10:19:38	c3nfqf
2017-12-28	2017-12-28 08:51:12	c3nfw2
2018-02-08	2018-02-08 07:38:35	c3nfu3
2018-01-10	2018-01-10 18:01:13	c3ncx0
2018-10-10	2018-10-10 16:10:56	c3nf5h
2017-06-24	2017-06-24 17:58:50	c3nfhg
2016-12-16	2016-12-16 18:07:35	c3nfkt
2018-06-10	2018-06-10 10:58:26	c3nfjr
2018-02-26	2018-02-26 16:35:18	c3nfmg
2018-04-10	2018-04-10 18:59:54	c3nfj4
2018-03-16	2018-03-16 15:11:26	c3nfj0

Cyclical Features



Hour: 0,1,2,3…23

Month: 1,2,3,4…12

In [6]:
```
import numpy as np

df['hr_sin'] = np.sin(df.hr*(2.*np.pi/24))
df['hr_cos'] = np.cos(df.hr*(2.*np.pi/24))
df['mnth_sin'] = np.sin((df.mnth-1)*(2.*np.pi/12))
df['mnth_cos'] = np.cos((df.mnth-1)*(2.*np.pi/12))
```

Source: http://blog.davidkaleko.com/feature-engineering-cyclical-features.html

Features

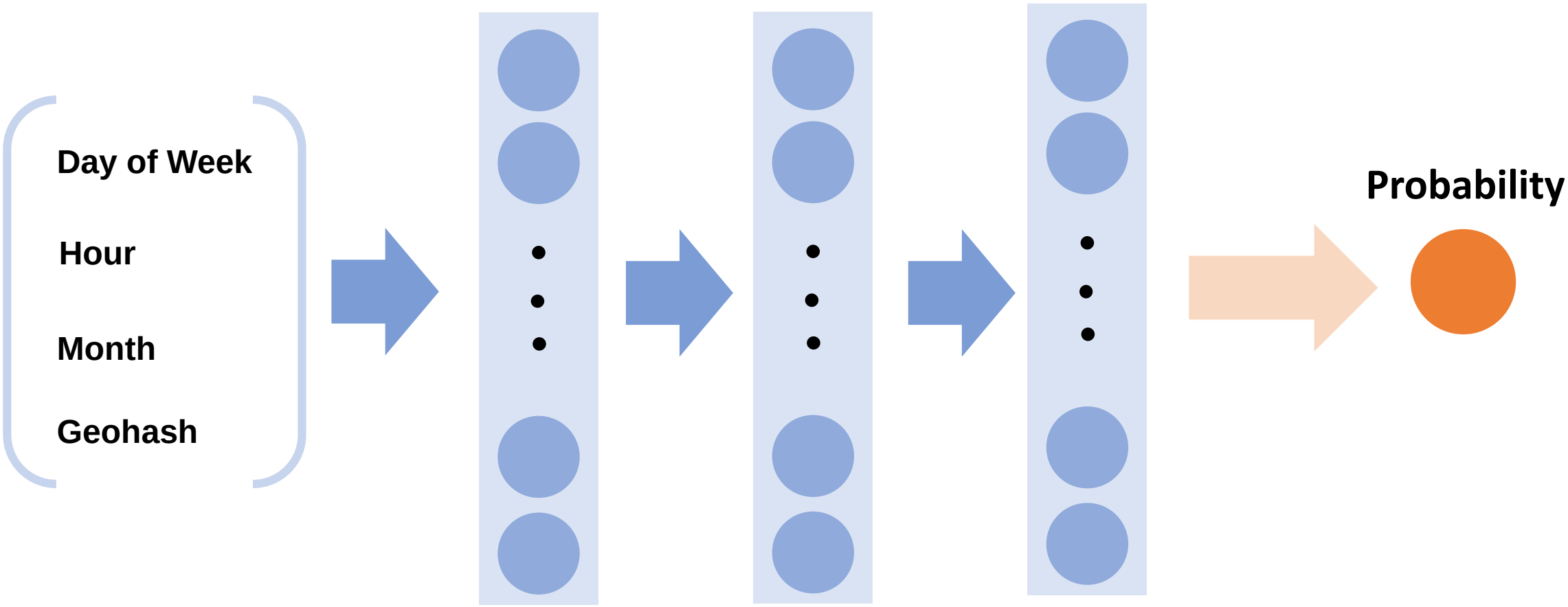
```
In [25]: all_data.head()
```

Out[25]:

	y	month_sin	month_cos	day_sin	day_cos	hour_sin	hour_cos	92nbhb	c3nc7z	c3nc90	...	c3ngp0	c3ngp1	c3ngp2	c3ngp8	c3ngpb
2017-11-15	1	-5.000000e-01	8.660254e-01	0.433884	-0.900969	-1.000000	-1.836970e-16	0	0	0	...	0	0	0	0	0
2018-09-01	1	-1.000000e+00	-1.836970e-16	-0.781831	0.623490	0.258819	-9.659258e-01	0	0	0	...	0	0	0	0	0
2017-12-28	1	-2.449294e-16	1.000000e+00	-0.433884	-0.900969	0.707107	-7.071068e-01	0	0	0	...	0	0	0	0	0
2018-02-08	1	8.660254e-01	5.000000e-01	-0.433884	-0.900969	0.866025	-5.000000e-01	0	0	0	...	0	0	0	0	0
2018-01-10	1	5.000000e-01	8.660254e-01	0.433884	-0.900969	-0.965926	2.588190e-01	0	0	0	...	0	0	0	0	0

5 rows × 969 columns

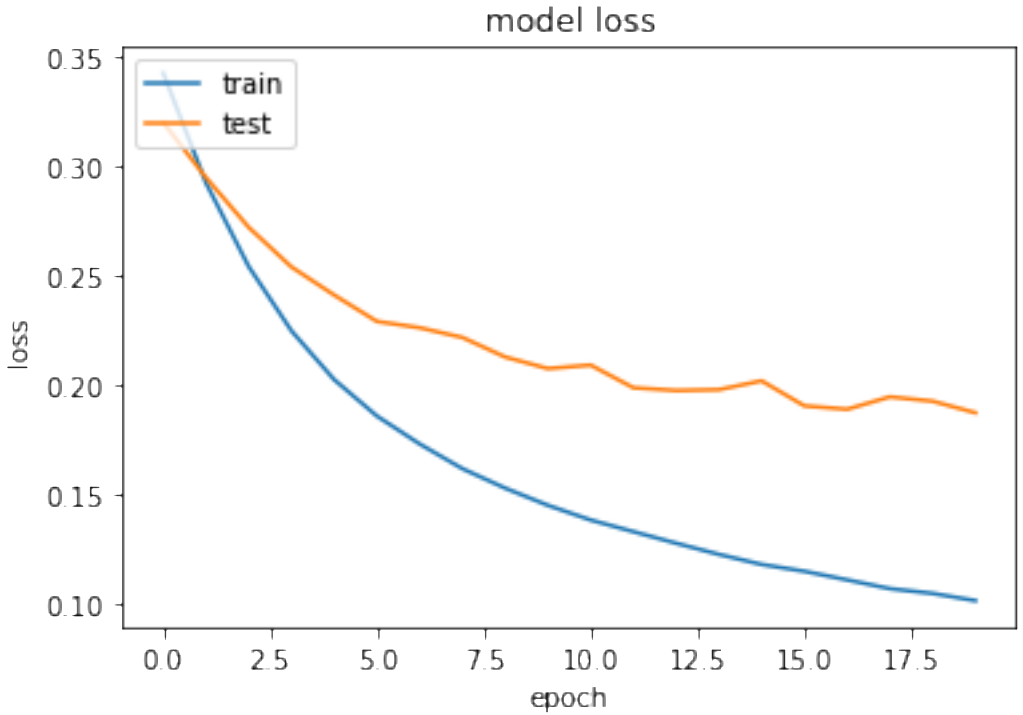
Building the Model



Training the Model

Training Data
(90%)

Test Data
(10%)



Front End Development

- **React** – fast development, deployment
- **Mapbox** – highly customizable map layers
- **Github** – free hosting for React web apps, fast prototyping