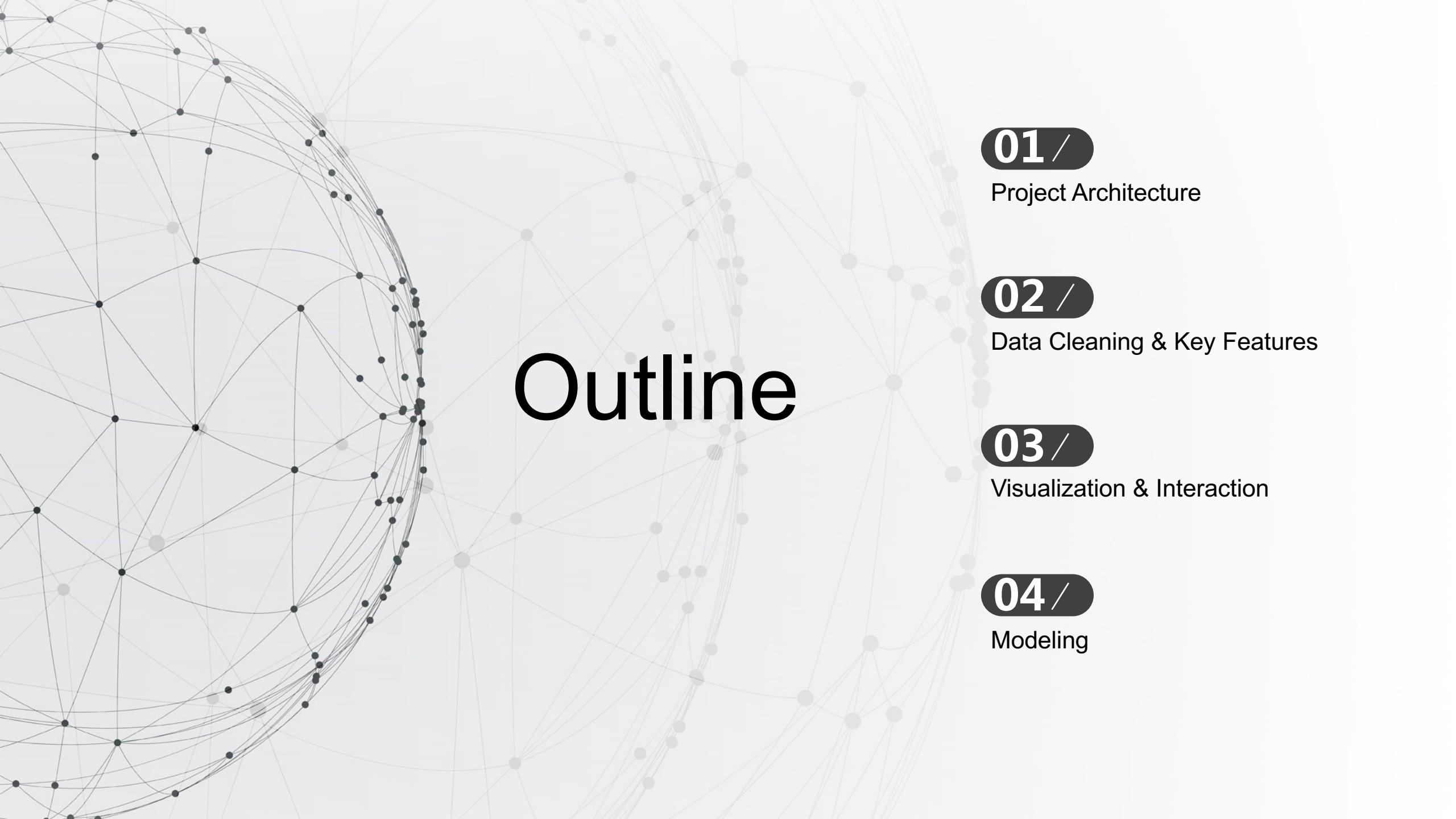


Crimes in Boston Prediction

Wei Jiang
Ci Chu
Kaikang Zhu

A faint, abstract network graph serves as the background for the slide. It consists of numerous small, dark gray dots representing nodes, connected by thin, light gray lines representing edges. The graph is more dense on the left side and becomes more sparse towards the right.

Outline

01 /

Project Architecture

02 /

Data Cleaning & Key Features

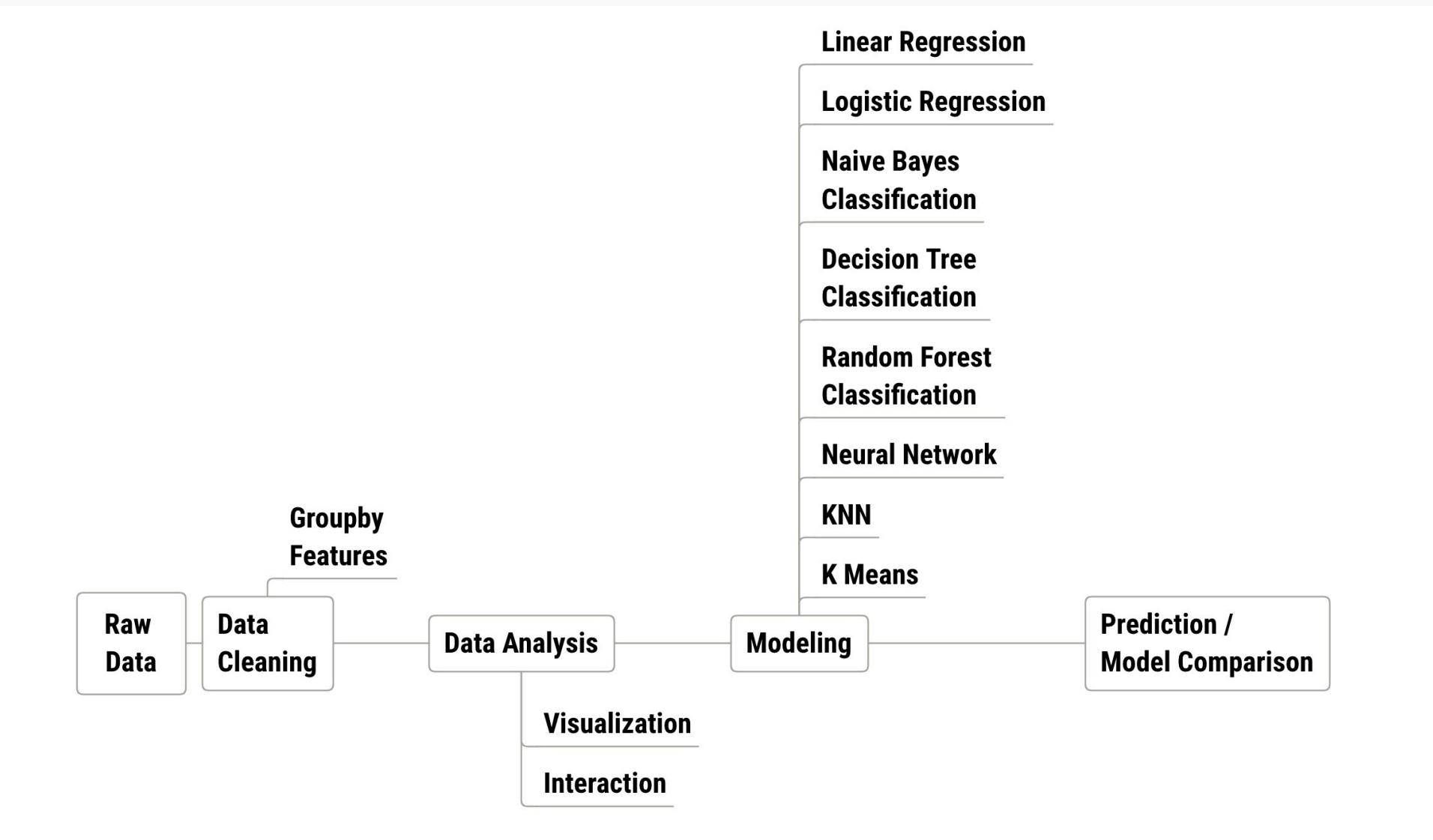
03 /

Visualization & Interaction

04 /

Modeling

Project Architecture

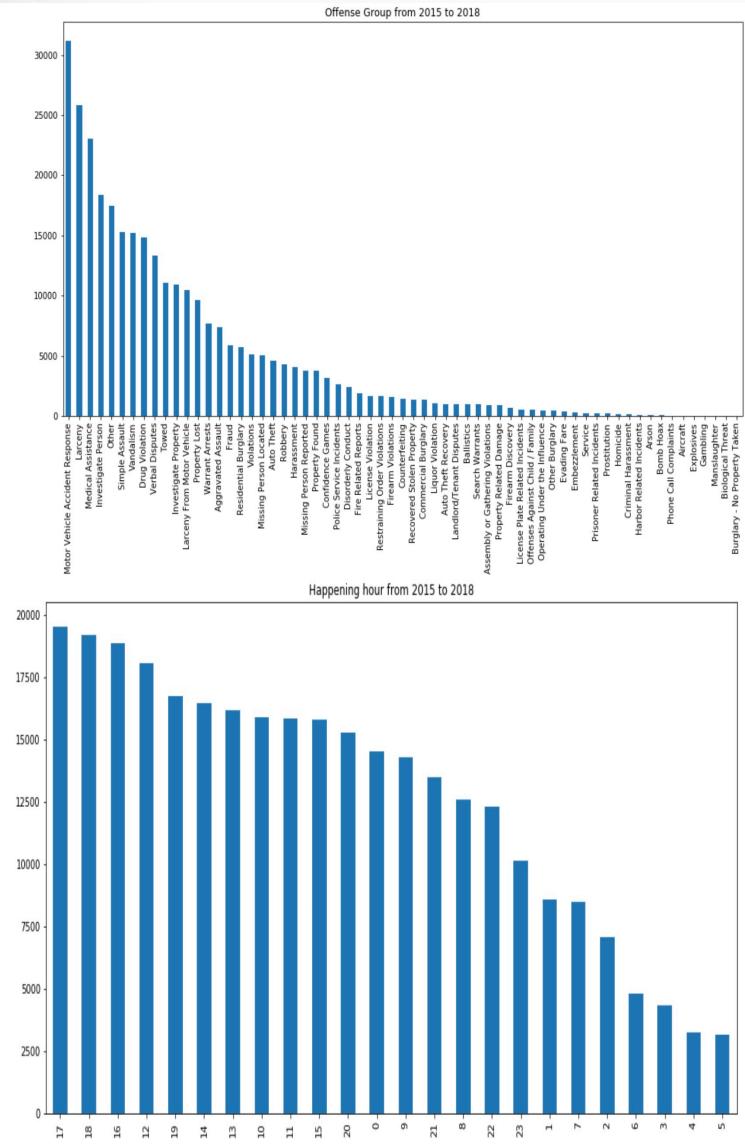


Data Cleaning & Key Features

	reportingArea	shooting	hour	day	month	year	lat	long
199132	618.0	0	8	7	8	2016	42.342279	-71.091522
265528	336.0	0	1	5	11	2015	42.305197	-71.068368
130047	342.0	0	16	3	4	2017	42.303419	-71.069557
169446	216.0	0	15	5	12	2016	42.332659	-71.051416
255577	149.0	0	12	1	11	2015	42.345979	-71.076712
211981	270.0	0	12	5	6	2016	42.333951	-71.075389

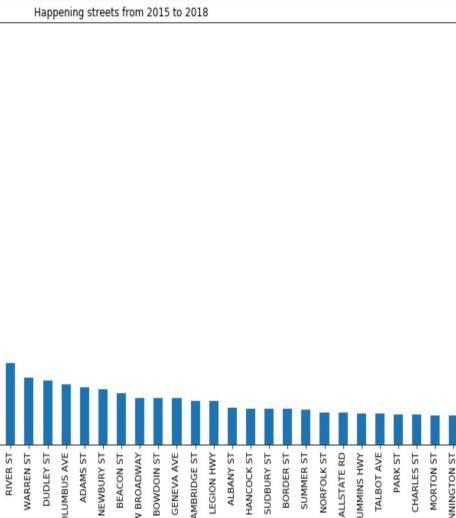
- **Reporting area:** Areas where crimes happened, preprocessed with pandas.to_numeric and one hot encoder
- **Shooting:** whether crimes contain shooting or not
- **Hour, Day, Month, Year:** Specific time for crimes
- **Lat, Long:** Latitude and Longitude for crime cases
- **Offense Group:** Group which crimes belong to, ex: Motor Vehicle Accident Response, Larceny

Data Visualization & Interaction



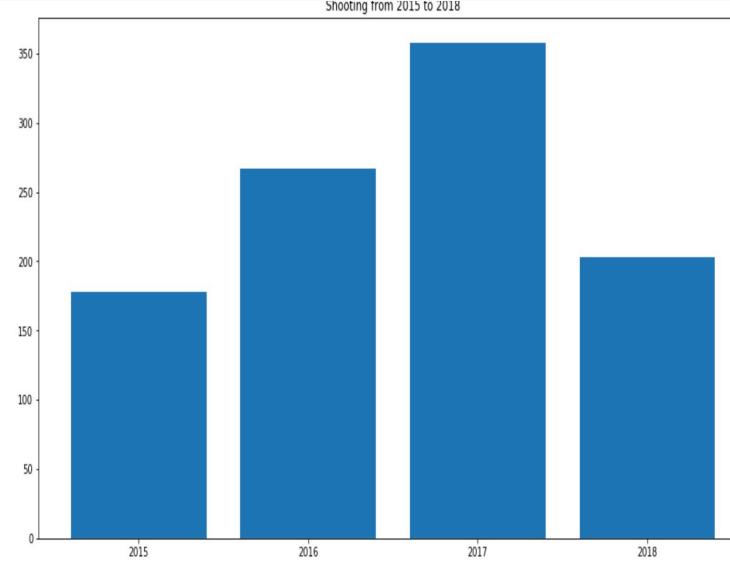
Offense Group bar chart

Head OG:
Motor Vehicle
Accident
Response,
Larceny



Happen Streets bar chart

Head HS:
Washington St,
Blue hill ave



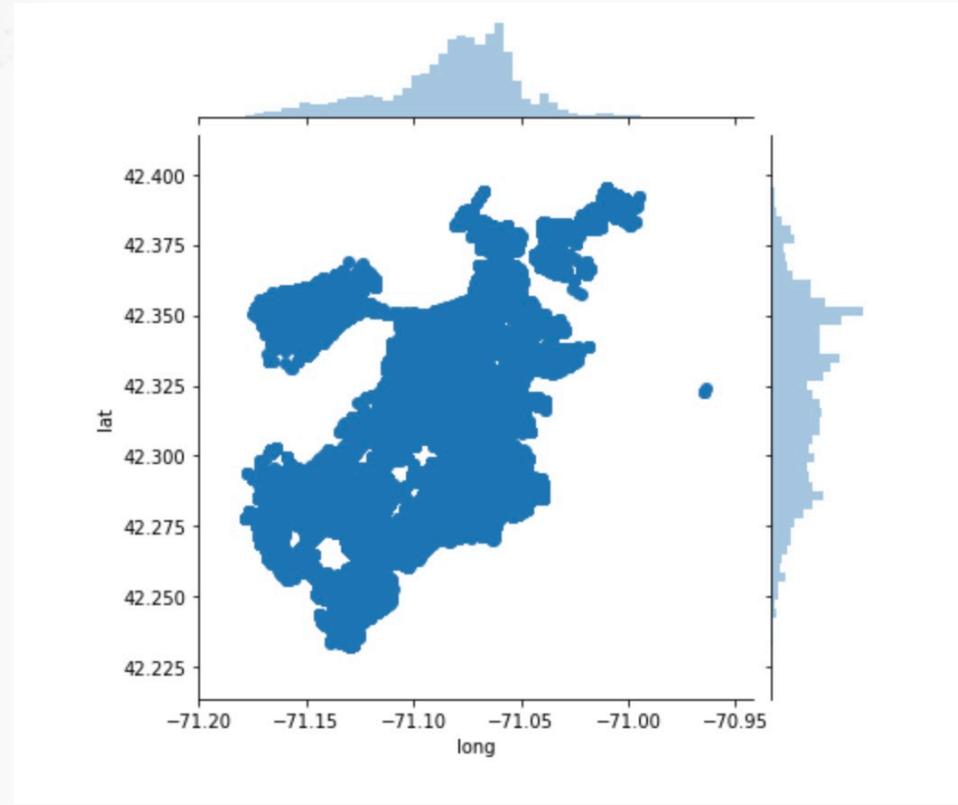
Happen Hour bar chart

Head HH:
17, 18, 16, 12

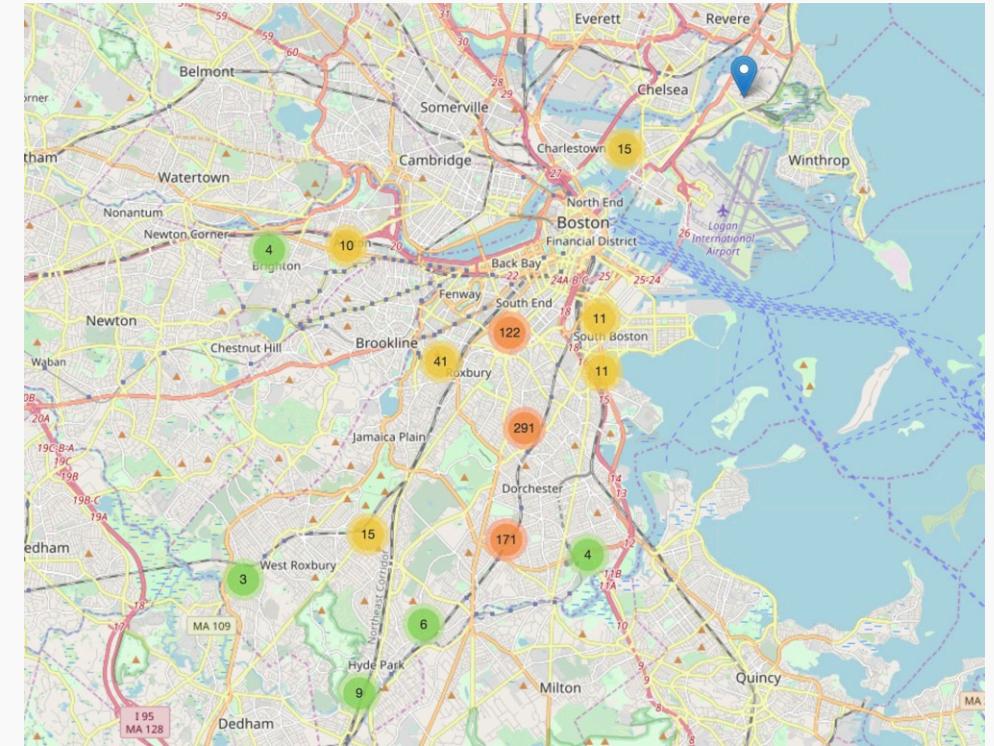
Shooting Year bar chart

Head SY:
2017, 2016

Data Visualization & Interaction



Crime spotting area **Geographical distribution**



Shooting area interaction page

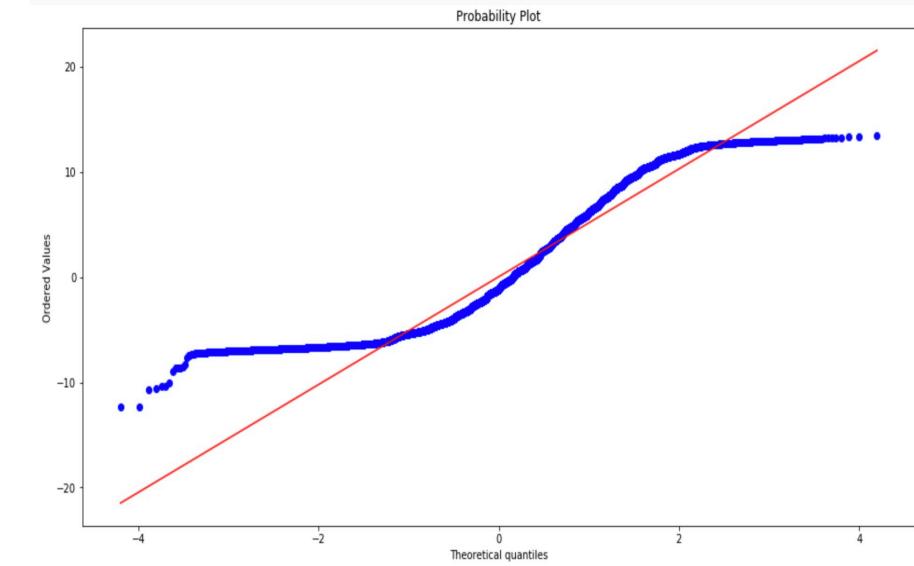
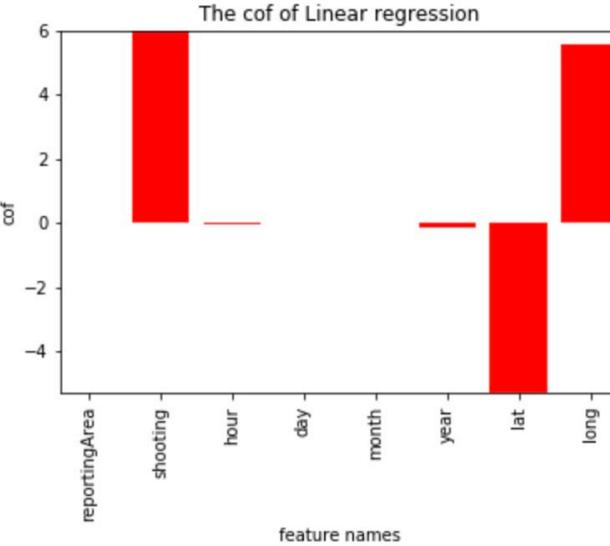
Modeling

- 1. Regression
 - 1.1 Linear Regression
 - 1.2 Logistic Regression

- 2. Classification
 - 2.1 Naïve Bayes Classification
 - 2.2 Decision Tree
 - 2.3 Random Forest

- 3. Neural Network

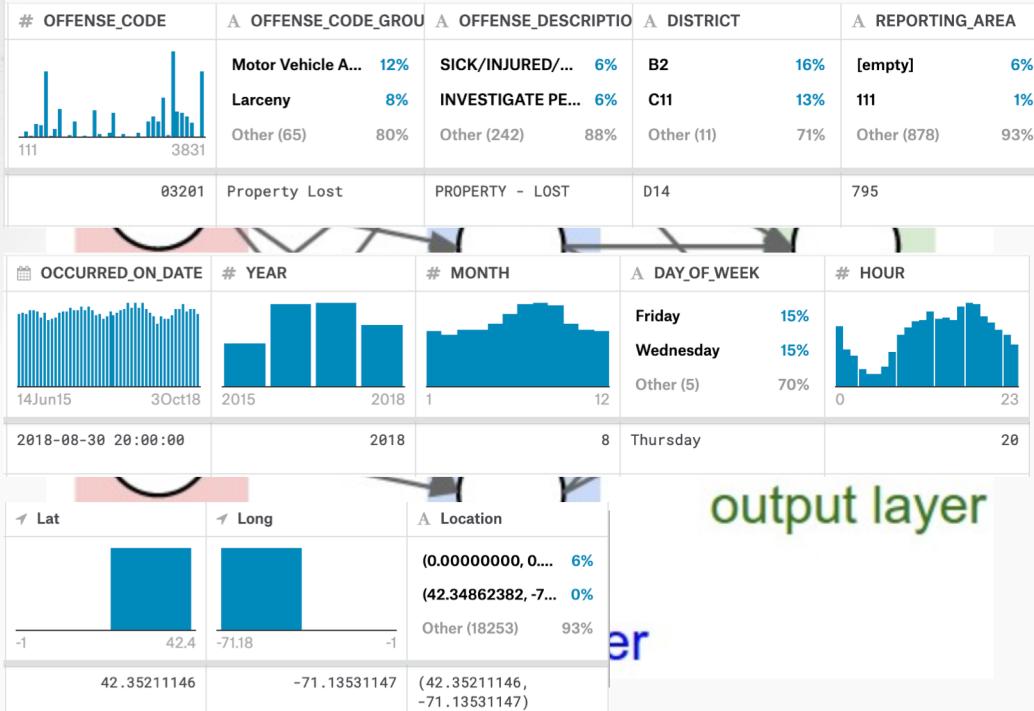
- 4. KNN & K-Means



	Logistic Regression	Gaussian NB	Bernoulli NB	Kneighbo rs Classifier	Decesion Tree Classifier	Random Forest Classifier
Accuracy	0.199	0.118	0.121	0.160	0.180	0.179
Precision	0.188	0.084	0.035	0.144	0.167	0.169
Recall	0.197	0.118	0.121	0.160	0.173	0.181
F1 score	0.153	0.050	0.029	0.138	0.169	0.168



Is it possible to use a neural network to predict crime's probability?



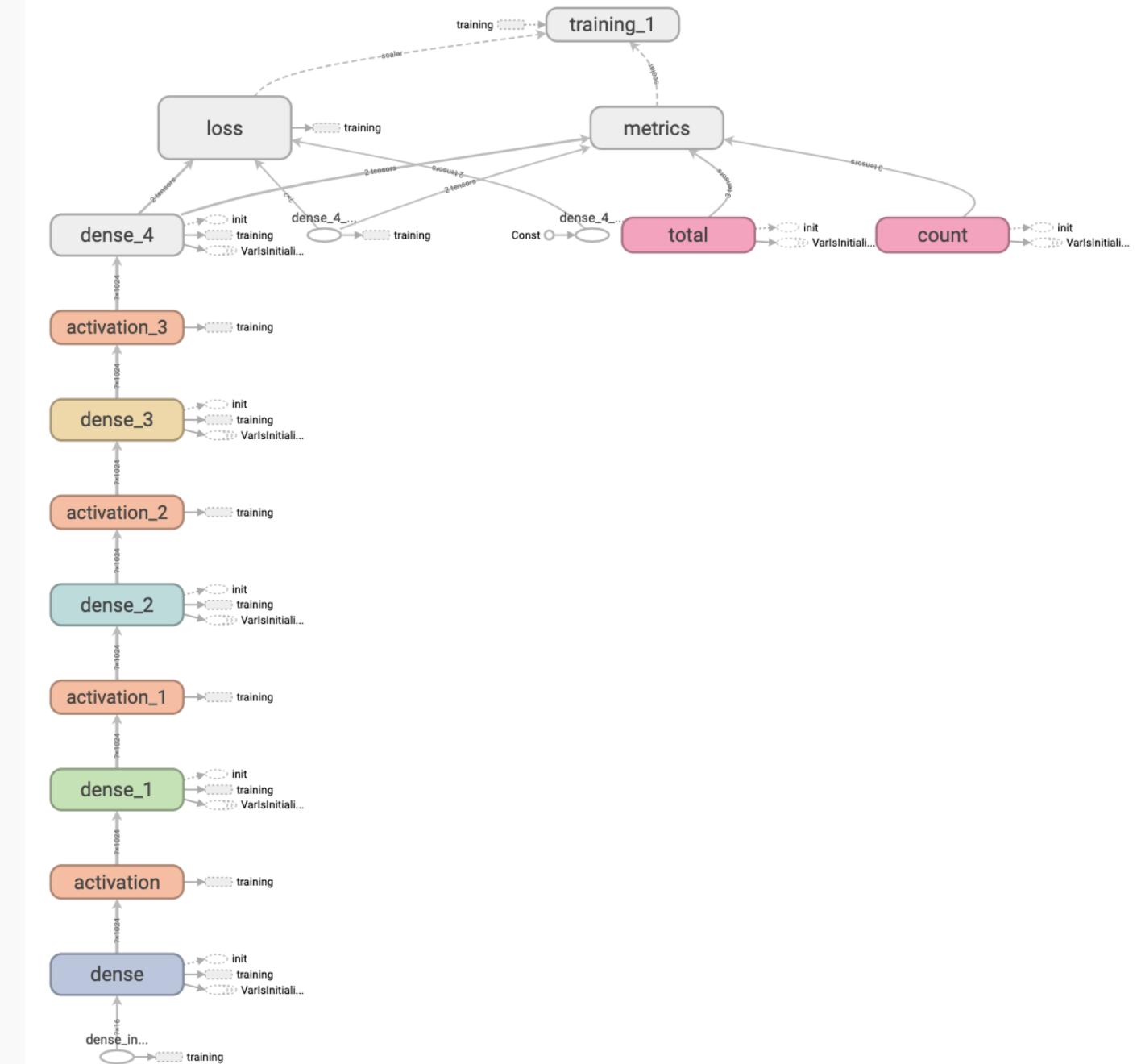
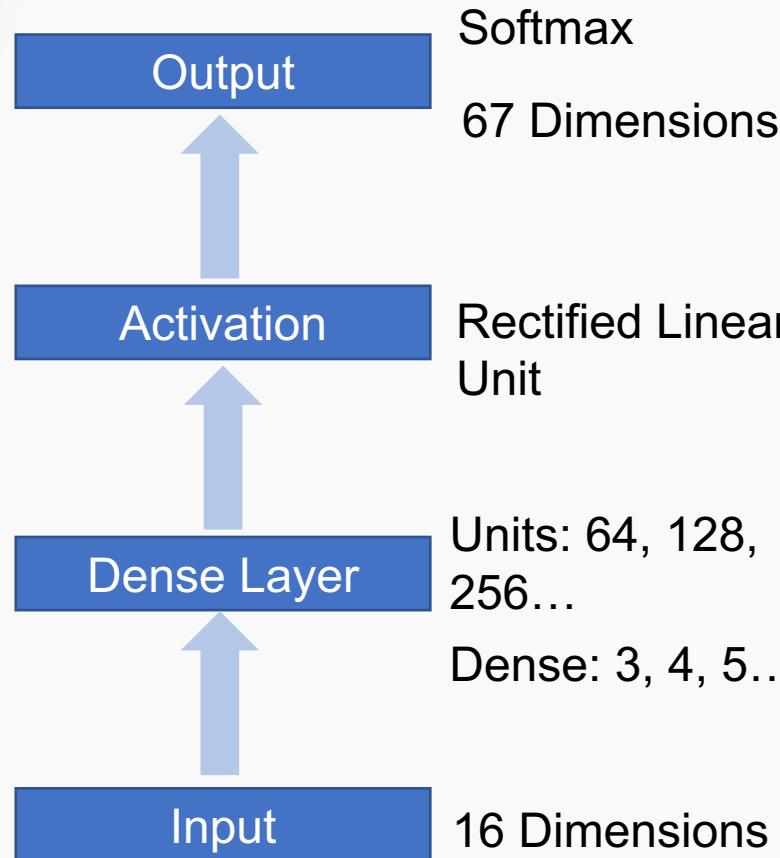
Input data:

'Year', 'Month', 'Day', 'Hour', 'Location'(District / Longitude, Latitude)

Output data:

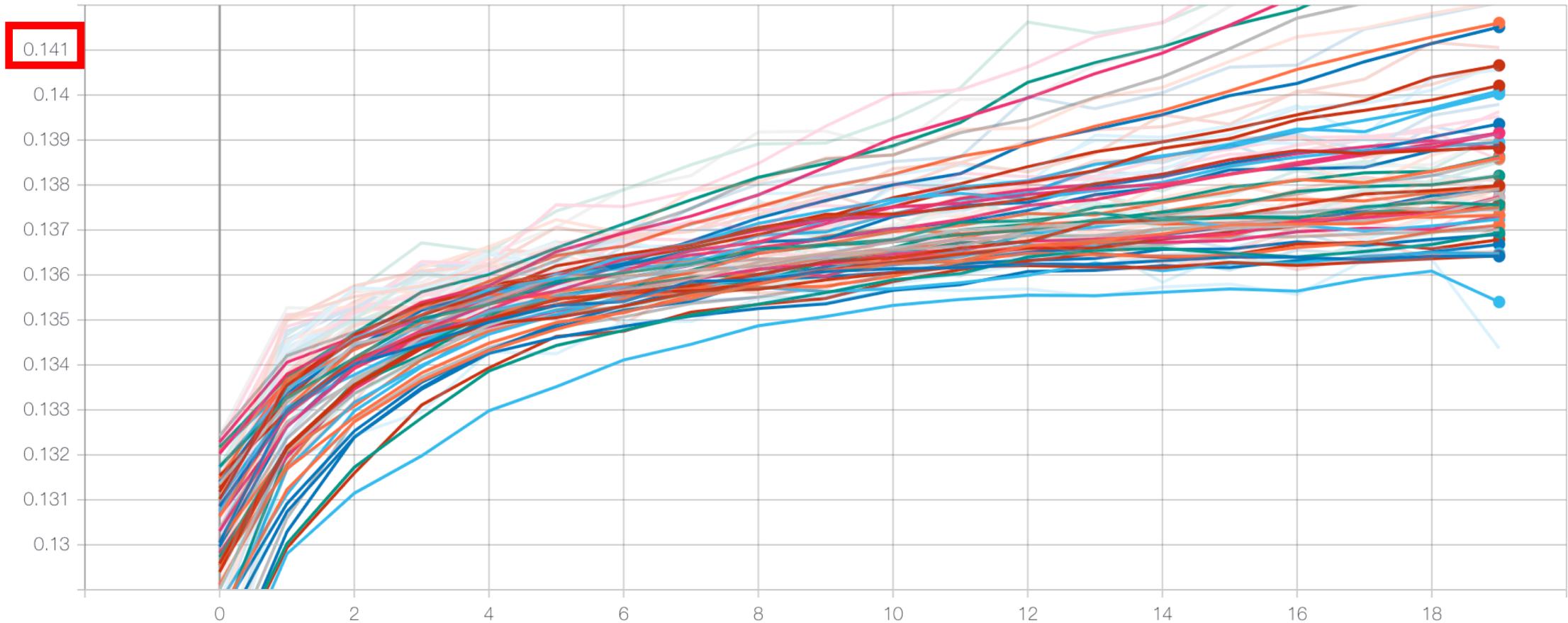
The probability of whether a kind of crime would happen

Basic Neural Network



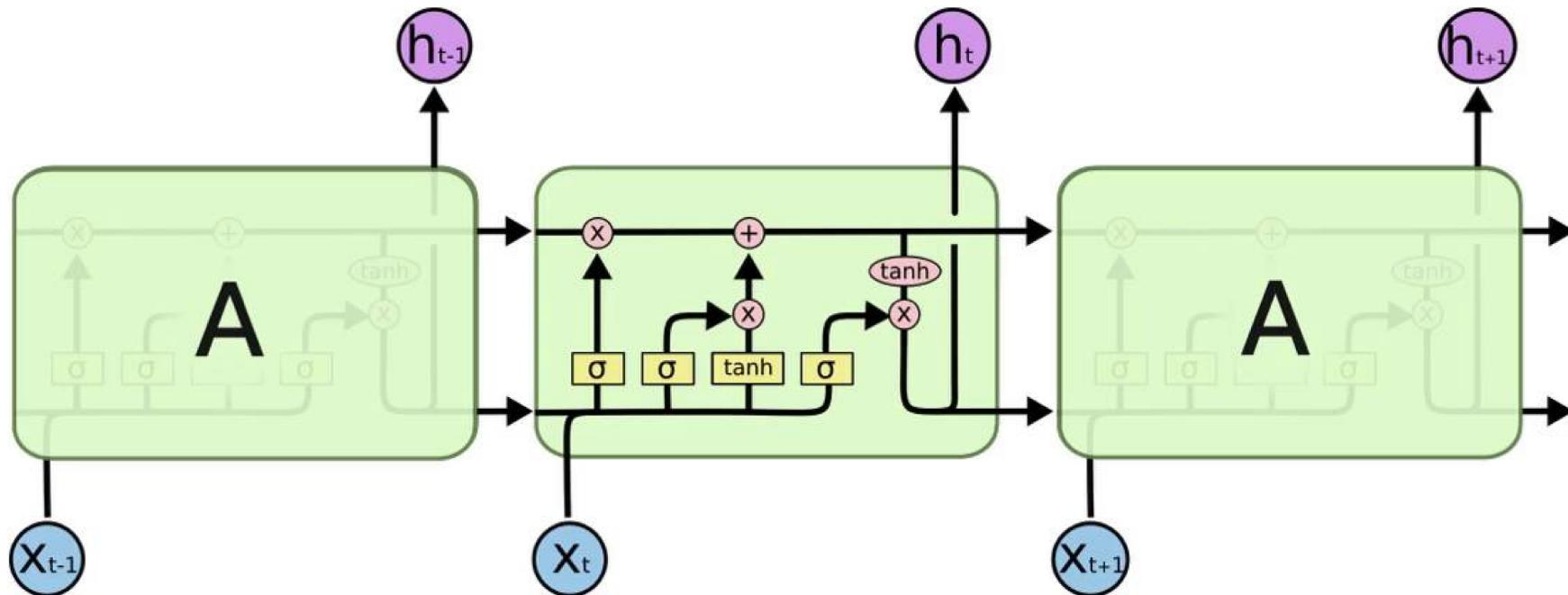
Accuracy and Loss

epoch_acc

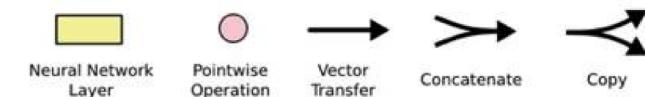


Recurrent Neural Network

Long-Short Term Memory module: LSTM



long-short term memory modules used in an RNN



Recurrent Neural Network

OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE	YEAR	MONTH	DAY_OF_WEEK	HOUR	UCR_PART	STREET	timestamp	count
DISTURBING THE PEACE	E18	495	NaN	2018-10-03 20:13:00	2018	10	Wednesday	20	Part Two	ARLINGTON ST	1538596800	1
PROPERTY - LOST	D14	795	NaN	2018-08-30 20:00:00	2018	8	Thursday	20	Part Three	ALLSTON ST	1535659200	1
THREATS TO DO BODILY HARM	B2	329	NaN	2018-10-03 19:20:00	2018	10	Wednesday	19	Part Two	DEVON ST	1538593200	1
ASSAULT - AGGRAVATED - BATTERY	A1	92	NaN	2018-10-03 20:00:00	2018	10	Wednesday	20	Part One	CAMBRIDGE ST	1538596800	1
AIRCRAFT INCIDENTS	A7	36	NaN	2018-10-03 20:49:00	2018	10	Wednesday	20	Part Three	PREScott ST	1538596800	1
VANDALISM	C11	351	NaN	2018-10-02 20:40:00	2018	10	Tuesday	20	Part Two	DORCHESTER AVE	1538510400	1

Recurrent Neural Network

```
Counter({1434326400: 14,  
1434330000: 10,  
1434333600: 7,  
1434337200: 3,  
1434340800: 1,  
1434344400: 1,  
1434348000: 2,  
1434351600: 8,  
1434355200: 12,  
1434358800: 14,  
1434362400: 15,  
1434366000: 10,  
1434369600: 17,  
1434373200: 17,  
1434376800: 8,  
1434380400: 16,  
1434384000: 13,  
1434387600: 16,  
1434391200: 8,  
1434394800: 17,})
```



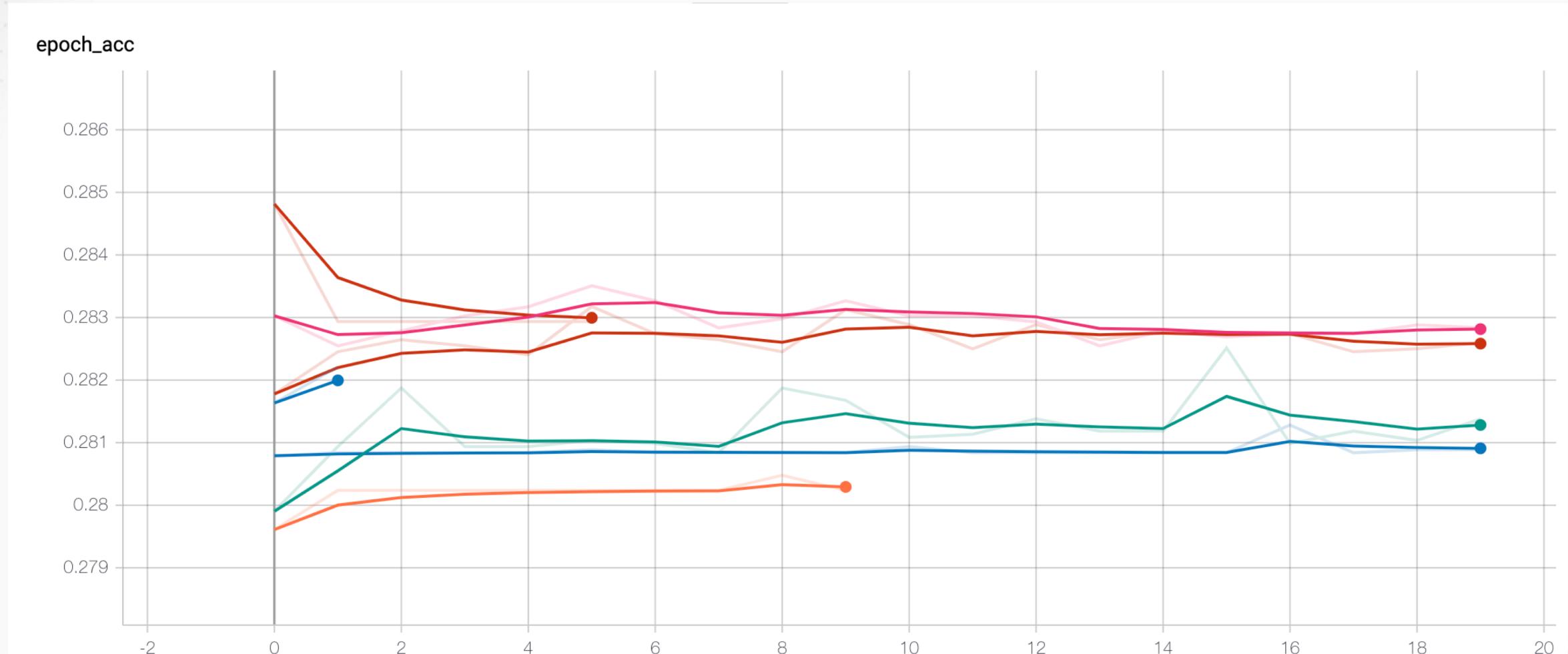
Input

72

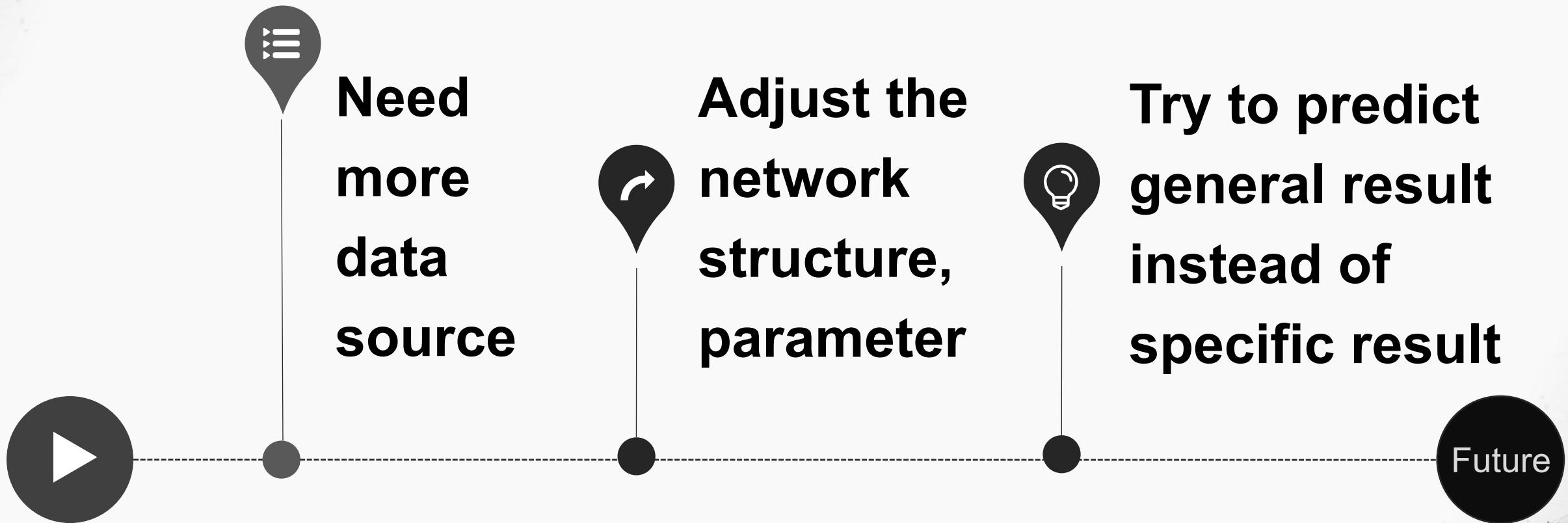
```
array([[14, 10, 7, ..., 9, 9, 10],  
[10, 7, 3, ..., 9, 10, 2],  
[ 7, 3, 1, ..., 10, 2, 3],  
...,  
[23, 16, 14, ..., 17, 9, 11],  
[16, 14, 12, ..., 9, 11, 10],  
[14, 12, 6, ..., 11, 10, 16]])
```

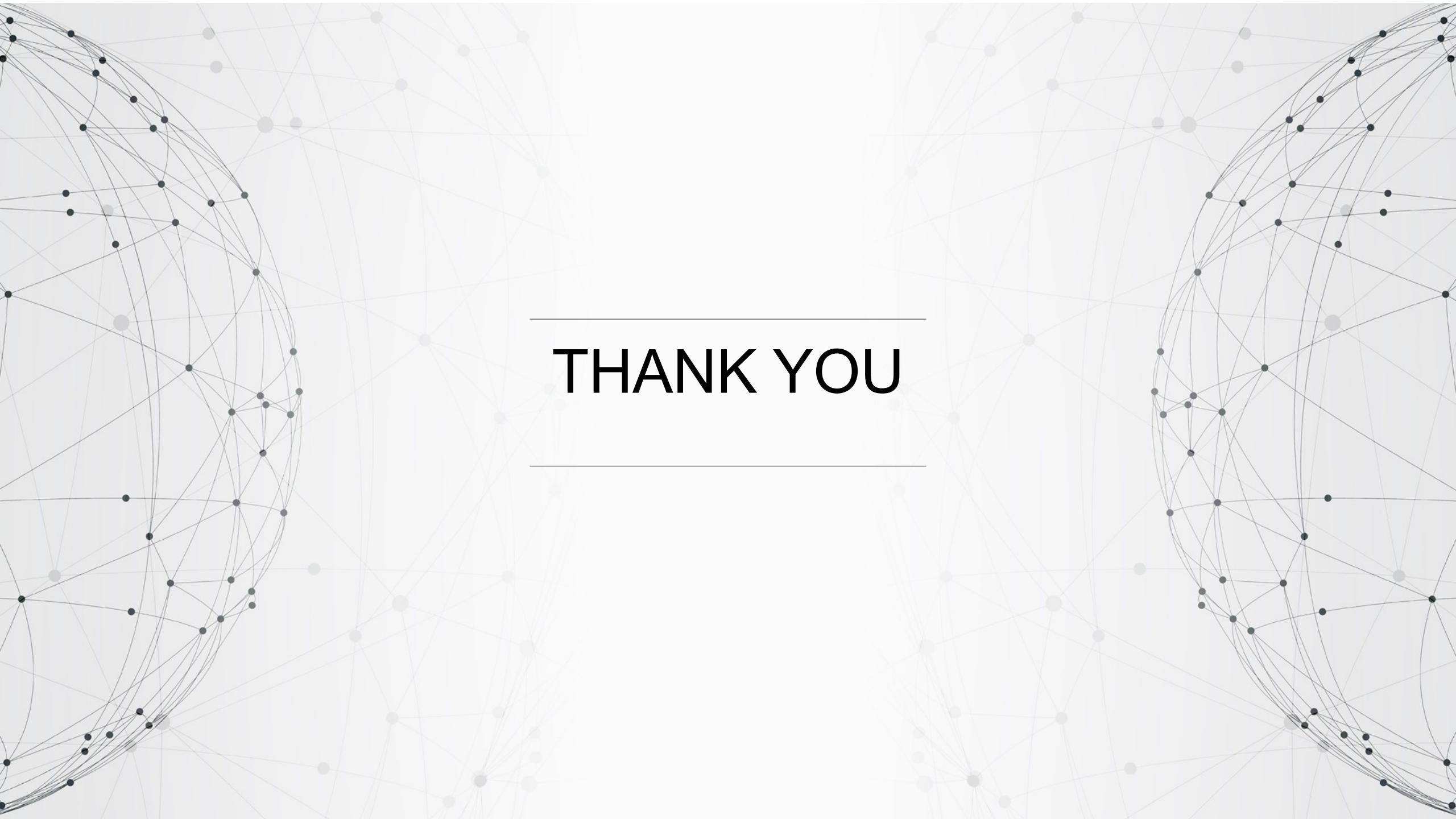
Output

Accuracy and Loss



Future work





THANK YOU
