

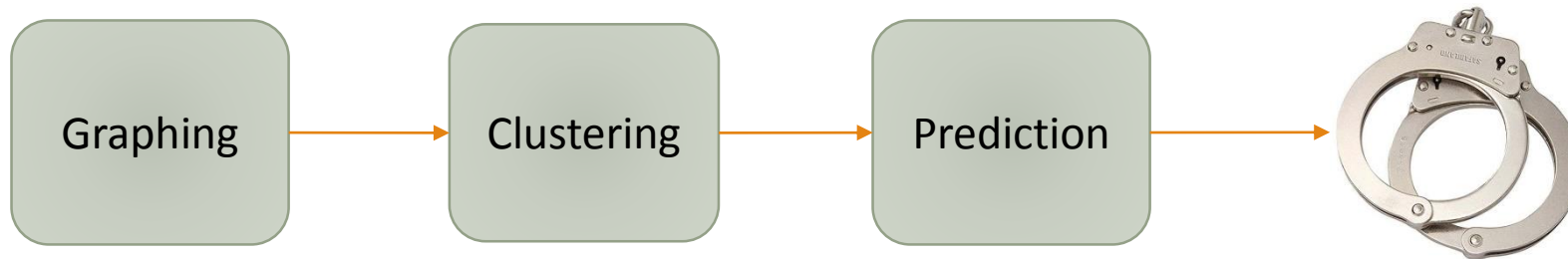
FIGHTING CRIME WITH BIG DATA



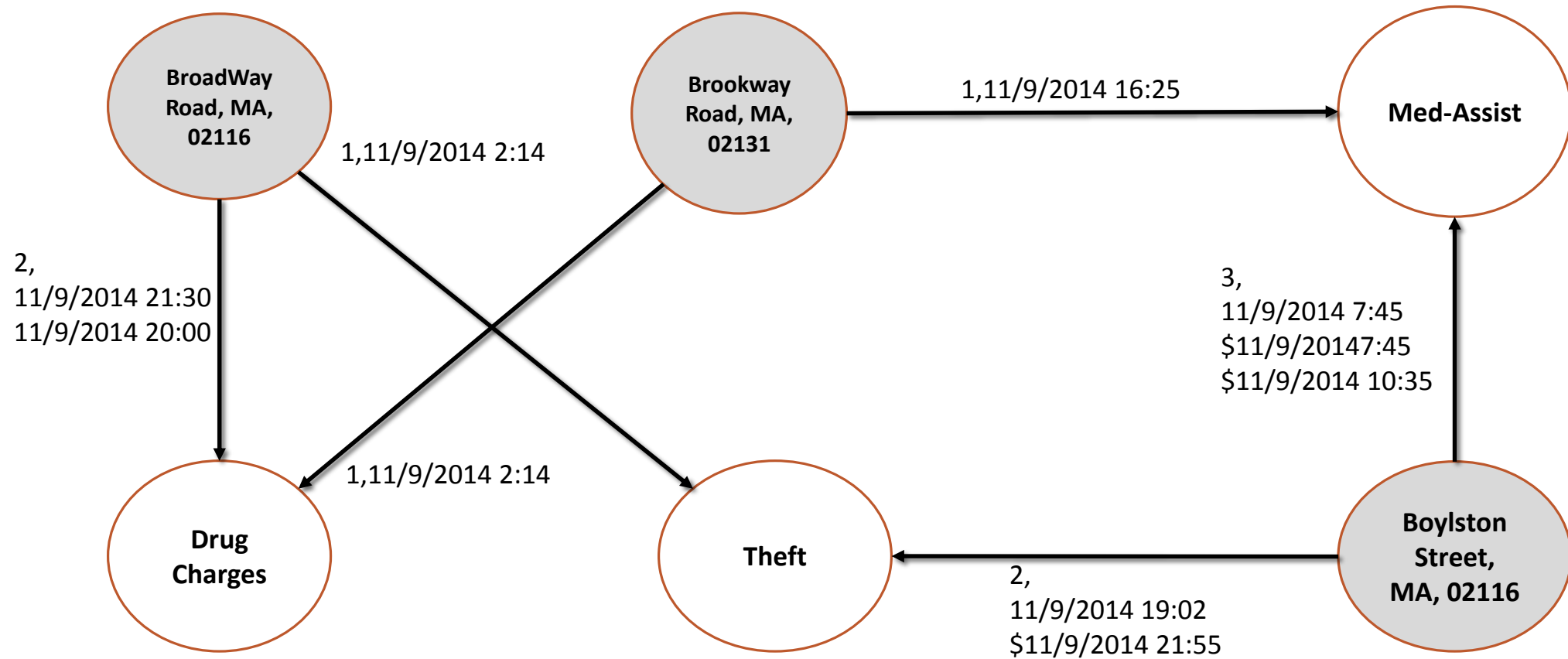
ABHIMANYU KULKARNI
NADIM MAHESANIYA
SHREYAS MAHIMKAR

Overview

- 1.5 million crimes per year.
- FBI and local police department have granular data about reported and unreported criminal activities.
- Can we leverage this data to prevent occurrence of crimes?



Graph Representation : Crime Data



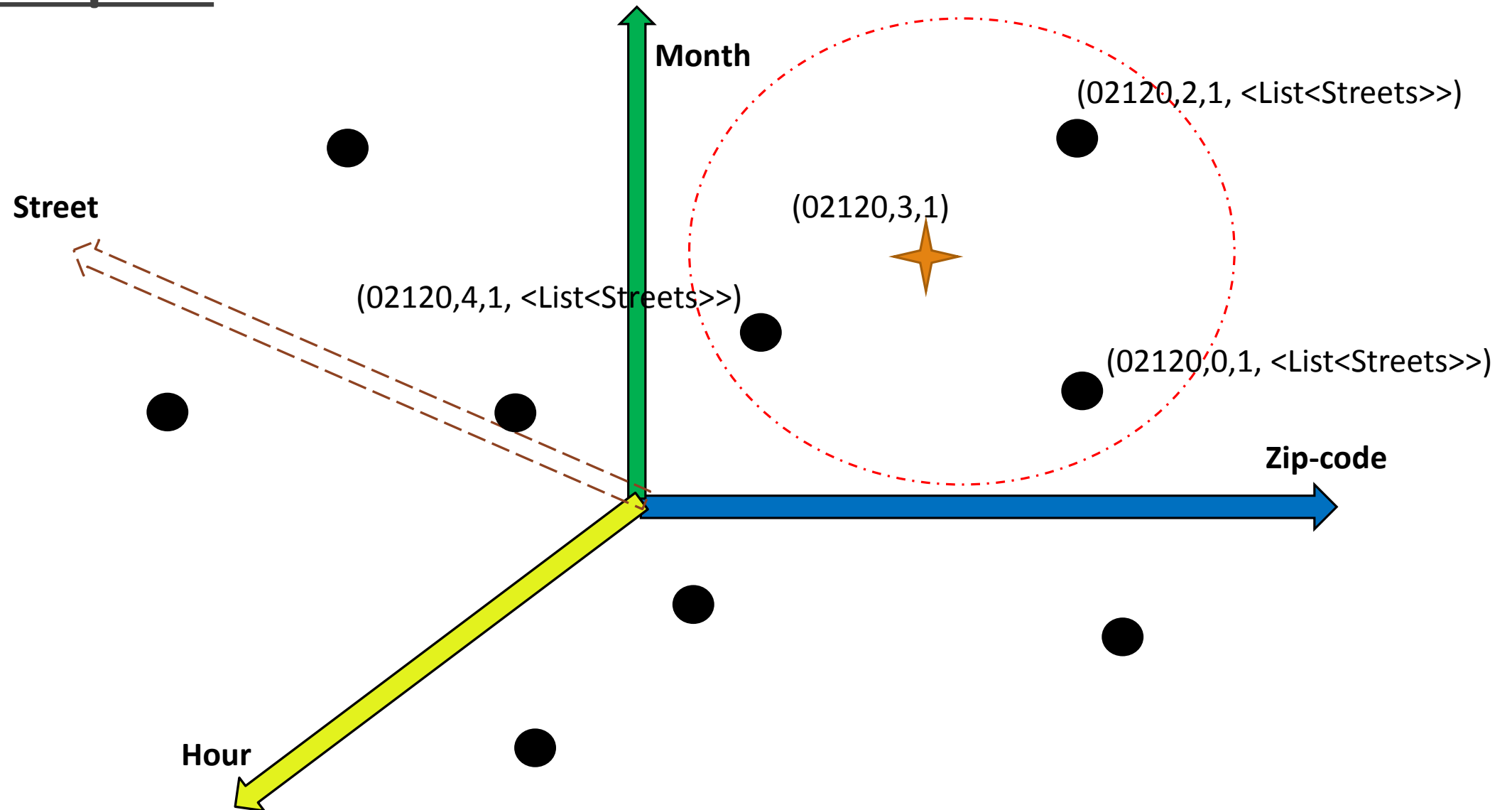
Clustering

- Clustering based on attributes zip-code , hour, month
- Weighting Schemes Followed:
 - Crime Occurred In Last 365 Days → 50
 - Crime Occurred At A Particular Hour → 75
 - Crime Occurred In A Particular Month → 20
- Formula for calculating weight of Street
$$\sum_{i=1}^n (i * \text{Recent} + i * \text{Hourly} + i * \text{Monthly})$$
- Example: 02115,23,11 → Street1:Weight1, Street2:Weight2 ...

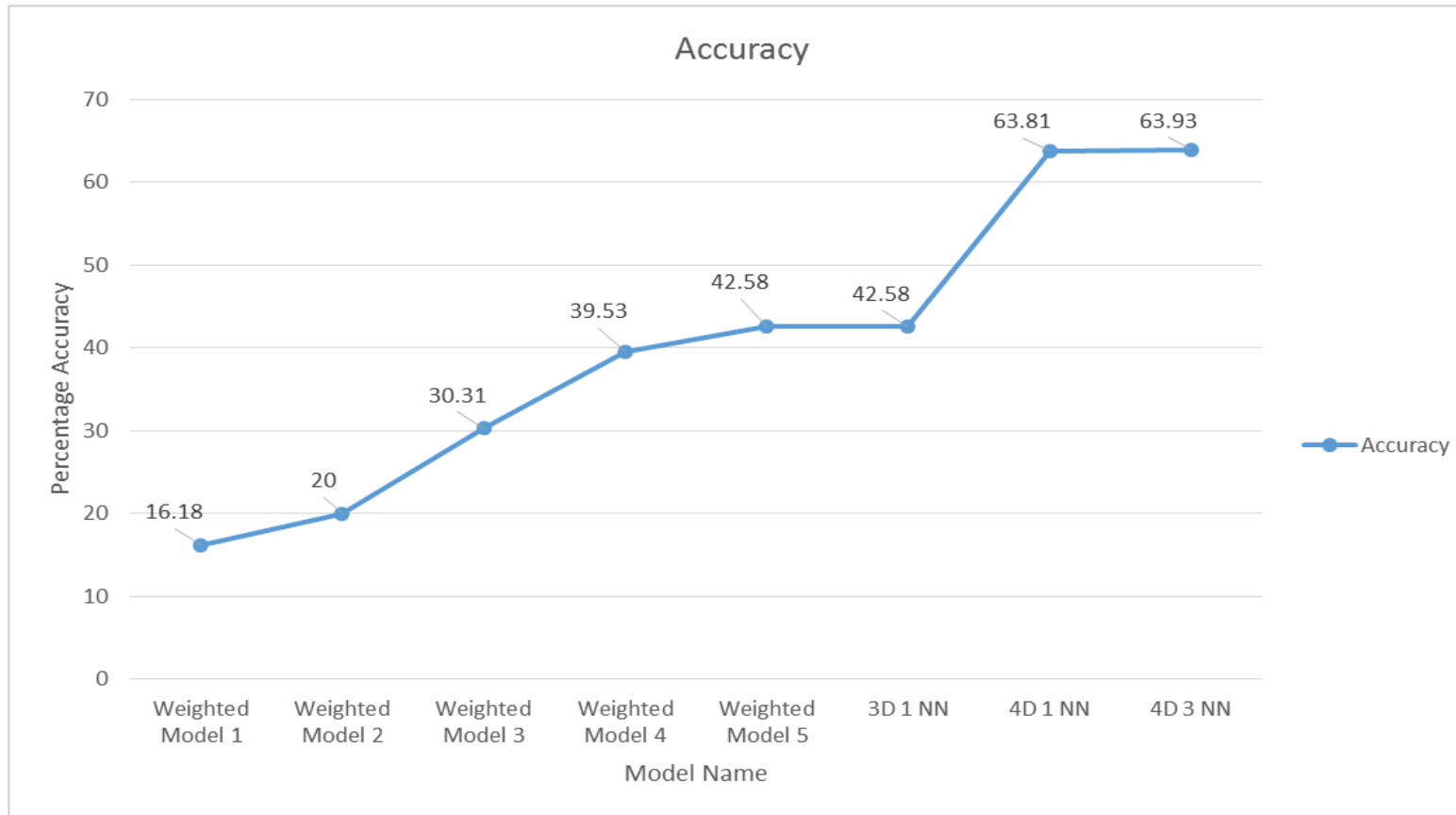
4 Dimensional Tree : 3 Nearest Neighbors

- Dimensions – (zip-code, hour, month, Street)
- Given a zip-code, hour and month, predict the streets where probability of occurrence of crime is high.
- Parallelism was achieved by partitioning based on value of zip-code.

Example:



Prediction Accuracy:



Model	M	R	H
WM1	5	12	-
WM2	50	25	-
WM3	50	25	-
WM4	20	50	-
WM5	20	50	75

Assistants

- Convert Flat Data To Graph → Adjacency List Representation.
- Storing Sample Data In RDBMS (~ 2 Million).
- Convert Latitude, Longitude To Zip-Code.
- Prediction Accuracy Calculator.

QUESTIONS?

