INTERNATIONAL TELEMATICS

INVESTIGATION OF LOCATION BASED SERVICES AND ANALYTICS

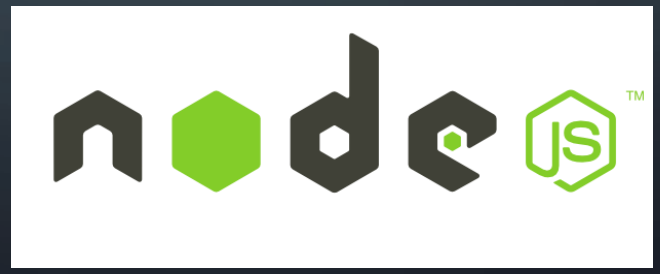
NIZAM SHAIK - 5695040

BTECH 451

FINAL PRESENTATION

THE PROJECT

- To develop an application to predict truck stop durations at pickup/delivery sites
- Use of previously recorded sample data
  - Truck details, Pickup/delivery site details, Currently recorded known stop duration of a truck at a site, etc
- Node.js and MongoDB

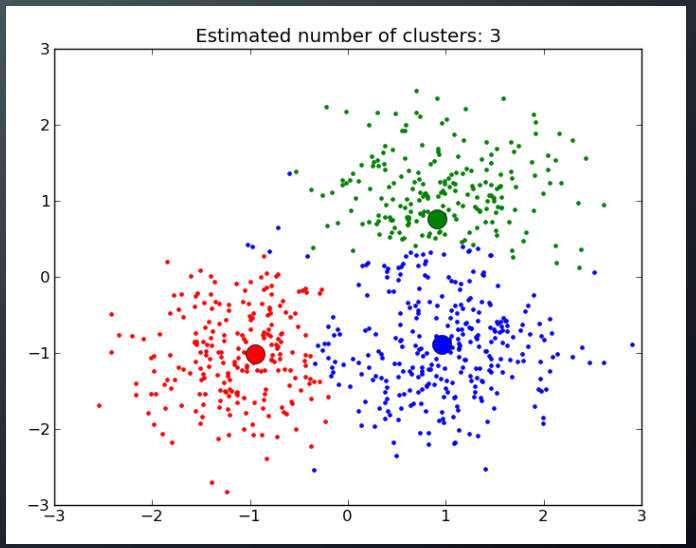


PREDICTION TECHNIQUES

- Data Mining Techniques
  - Supervised Learning
  - Unsupervised Learning
- Cluster Analysis (Unsupervised)
  - Unsure about the data structure
  - Possibility of no error/reward signals to evaluate a solution.
  - Adaptable to changes

CLUSTER ANALYSIS

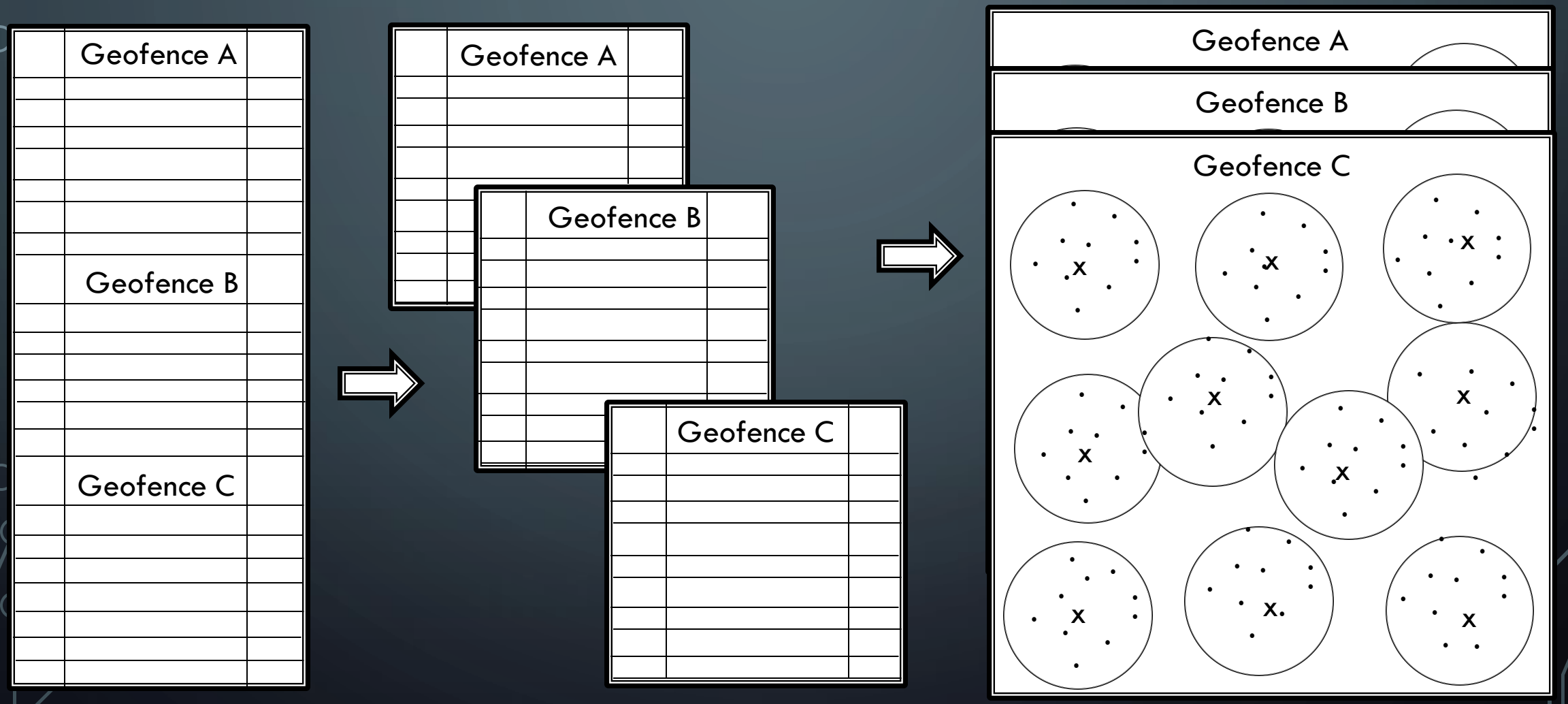
- Divides data into groups based on related information between data entries.
- K-Means Clustering Algorithm
  - Initially choose K number of total clusters/ centroids
  - Assign each data object to its nearest cluster/centroid



IMPLEMENTATION METHOD

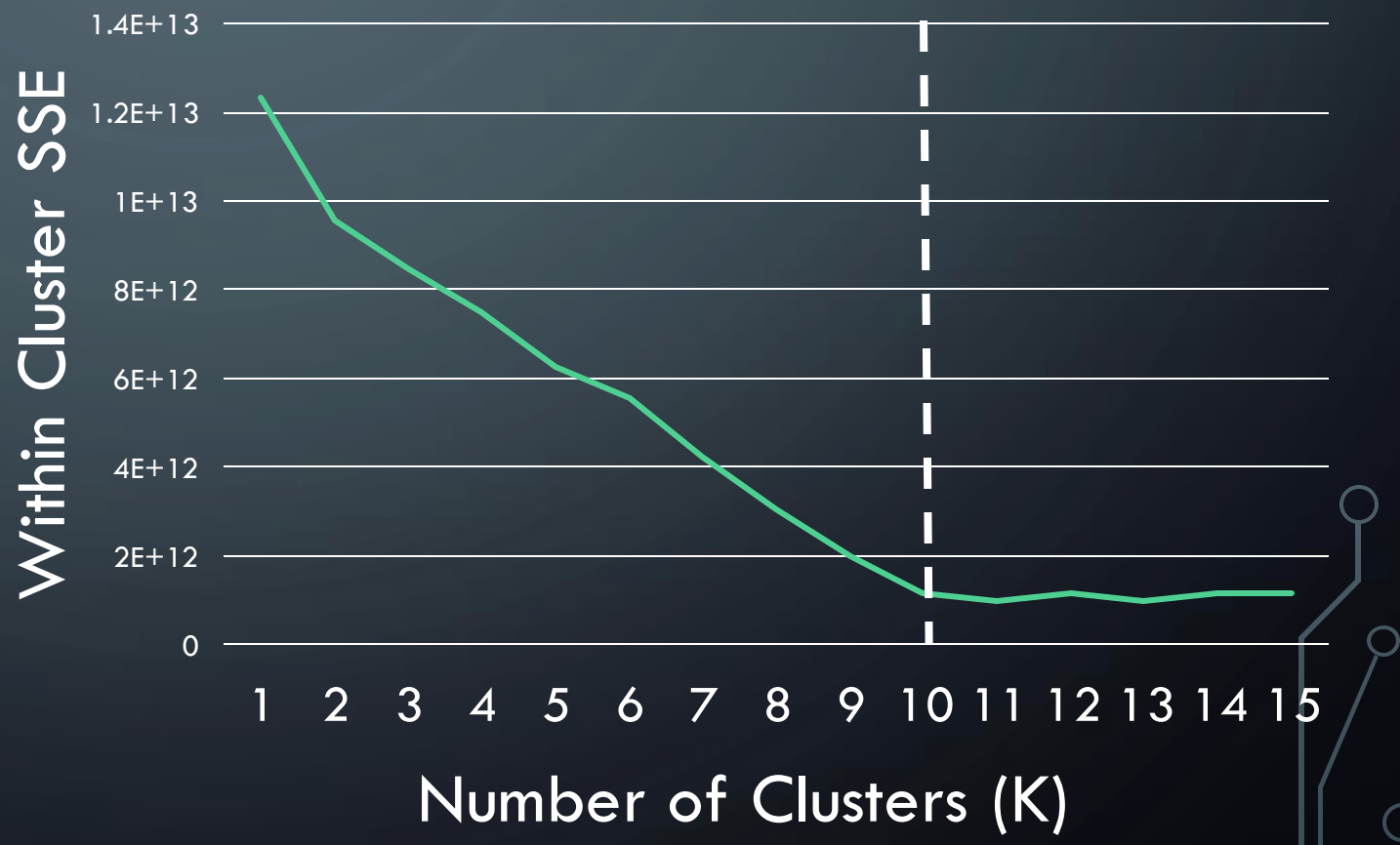
- Partition the data based on the different Geofences (locations of stops) recorded in the sample data
- Randomly select 10 points in each Geofence Partition – K number of initial centroids
- Assign the data objects in each partition to its nearest centroid to form clusters around the centroid
  - Nearest centroids found using Euclidean Distance

IMPLEMENTATION METHOD



NUMBER OF CLUSTERS

- Plot the sum of square errors within each cluster against K
- After “elbow point” further increasing K yields no improvements in the SSE



IMPLEMENTATION TECHNIQUES

- Nearest Centroid Measure
  - Euclidean Distance
    - Measure the distance of similarity of two data entries by comparing their attributes
- Evaluations
  - Internal Cluster Validation
    - Silhouette Measure (Average inter-cluster distance measure)
  - Final Prediction Evaluation
    - 10x10-Fold Cross Validation

EUCLIDEAN DISTANCE

- $d(A,B) = \sqrt{w_1(A_1 - B_1)^2 + w_2(A_2 - B_2)^2 \ldots + w_n(A_n - B_n)^2}$

Example - Data attributes are VehicleID, Month, and Duration (500s max)

- A = (“1001”, June, 400) and B = (“1001”, July, 150)

$\rightarrow \sqrt{0.25(1001 == 1001)^2 + 0.25(June == July)^2 + 0.5(0.8 - 0.3)^2}$

$\rightarrow \sqrt{0.25(0)^2 + 0.25(1)^2 + 0.5(0.5)^2}$

$\rightarrow \sqrt{0 + 0.25 + 0.125}$

$\rightarrow d(A,B) = 0.61$ = Distance between data entry A and data entry B

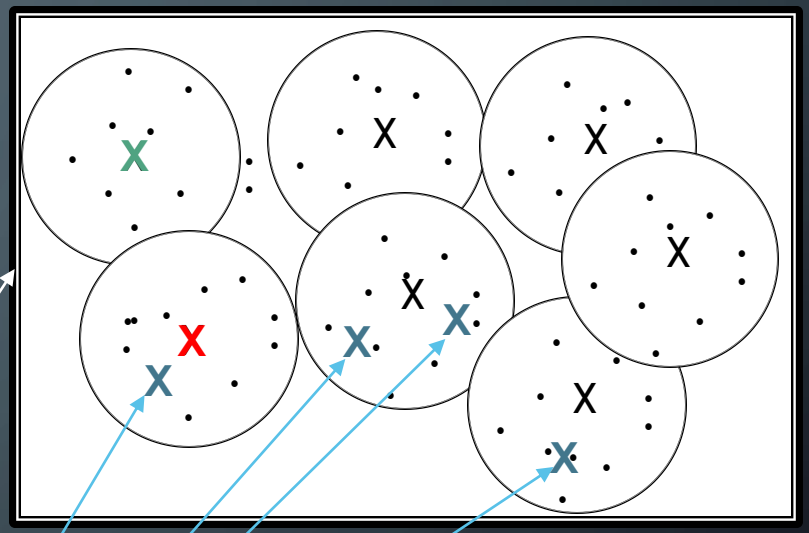
INTERNAL CLUSTER VALIDATION

- Silhouette Measure
  - $s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}}, -1 \leq s(i) \leq 1$
  - a(i) = The average distance of i to all other data entries in its cluster

    b(i) = The distance between i and closest neighbouring cluster
- A positive value of s(i) means that the data entry is appropriately clustered, where a negative values means it is not

FINAL PREDICTION EVALUATION

- 10x10 – Fold Cross Validation



- Data entry X's stop duration will be predicted with centroid X's stop duration
- The actual recorded stop duration for entry X compared with X and X
- If X-X is less than X-X then its predicted stop duration is the most accurate it can be

THE APPLICATION

RELATED WORK

- Supervised Vs Unsupervised Learning
  - Chaovalit, P., & Zhou, L. (2005, January). Movie review mining: A comparison between supervised and unsupervised classification approaches. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 2005. HICSS'05.* (pp. 112c-112c). IEEE.
- Clustering for Recommender Systems
  - Zhang, F., Liu, H., & Chao, J. (2010). A Two-stage Recommendation Algorithm Based on K-means Clustering In Mobile E-commerce. *Journal of Computational Information Systems*, *6*(10), 3327-3334.
  - Kim, T. H., & Yang, S. B. (2005). An effective recommendation algorithm for clustering-based recommender systems. In *AI 2005: Advances in Artificial Intelligence* (pp. 1150-1153). Springer Berlin Heidelberg.

RELATED APPLICATIONS

- Predictive Analytics for Traffic – Microsoft Research
  - Infer and predict the flow of traffic at different times in the future.
  - Machine learning tool which uses live streams and large amounts of historical data.
  - Interferences such as weather, major events, accidents etc.
- Bus Arrival Time Prediction Method for ITS Application
  - Stop duration at traffic lights cause the biggest errors in bus arrival time predictions.
  - Introduce prediction method which incorporates traffic light information.
  - Predict stop duration at traffic light giving more accurate bus arrival time prediction.
  - Son, Bongsoo, et al. "Bus arrival time prediction method for its application. "*Knowledge-Based Intelligent Information and Engineering Systems*. Springer Berlin Heidelberg, 2004.

THANK YOU