

ELYSIAN EDU
AI LEARNING CENTER



Transformation of DATA into Business INSIGHTS

CHANGE THE WORLD THROUGH
SCIENCE AND TECHNOLOGY

THET SU WIN

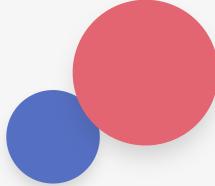


Here are the Outlines !!

The topic will cover these 4 main sessions:

- 1.What is DATA & Information & Insights
- 2.Predictive Modelling (Data Science)
- 3.Data Mining using Graph Representation
- 4.Case Study



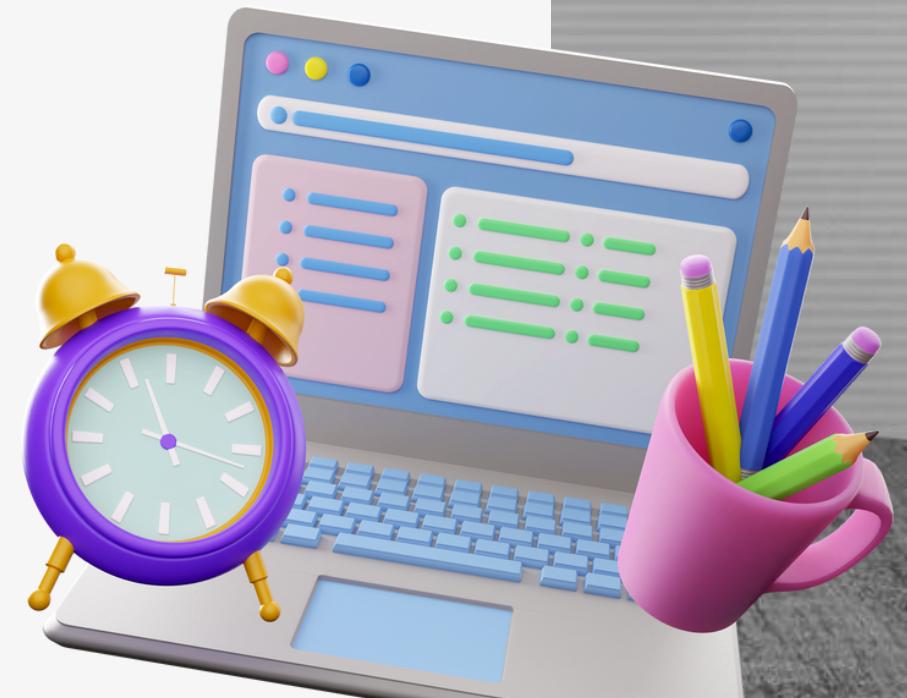


ABOUT ME

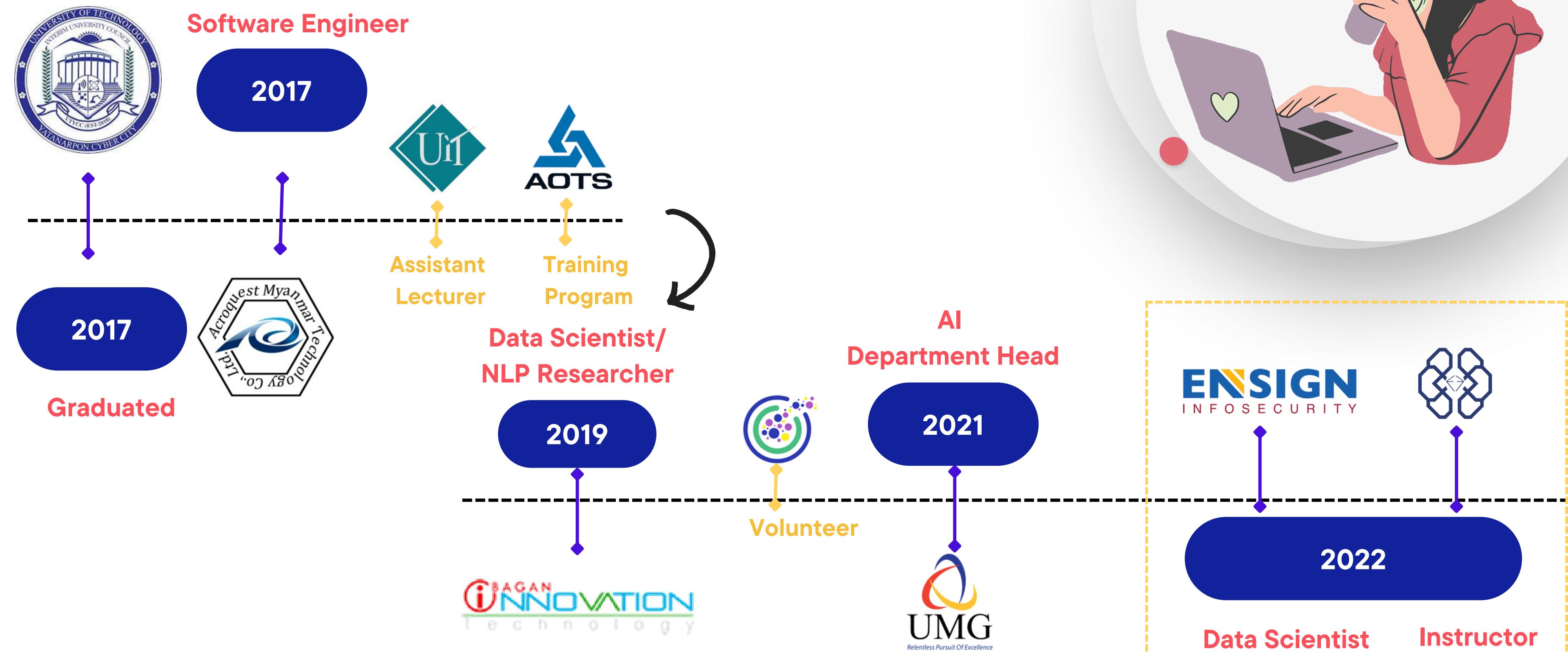
Thet Su Win

Data Scientist @ EnsignInfoSecurity, Singapore
Instructor @ Elysian AI Learning School

B.E (Information Science and Technology)
University of Technology (Yatanarpon Cyber City)



Work Experiences



1



DATA vs INFORMATION vs INSIGHTS

Let's get started.....

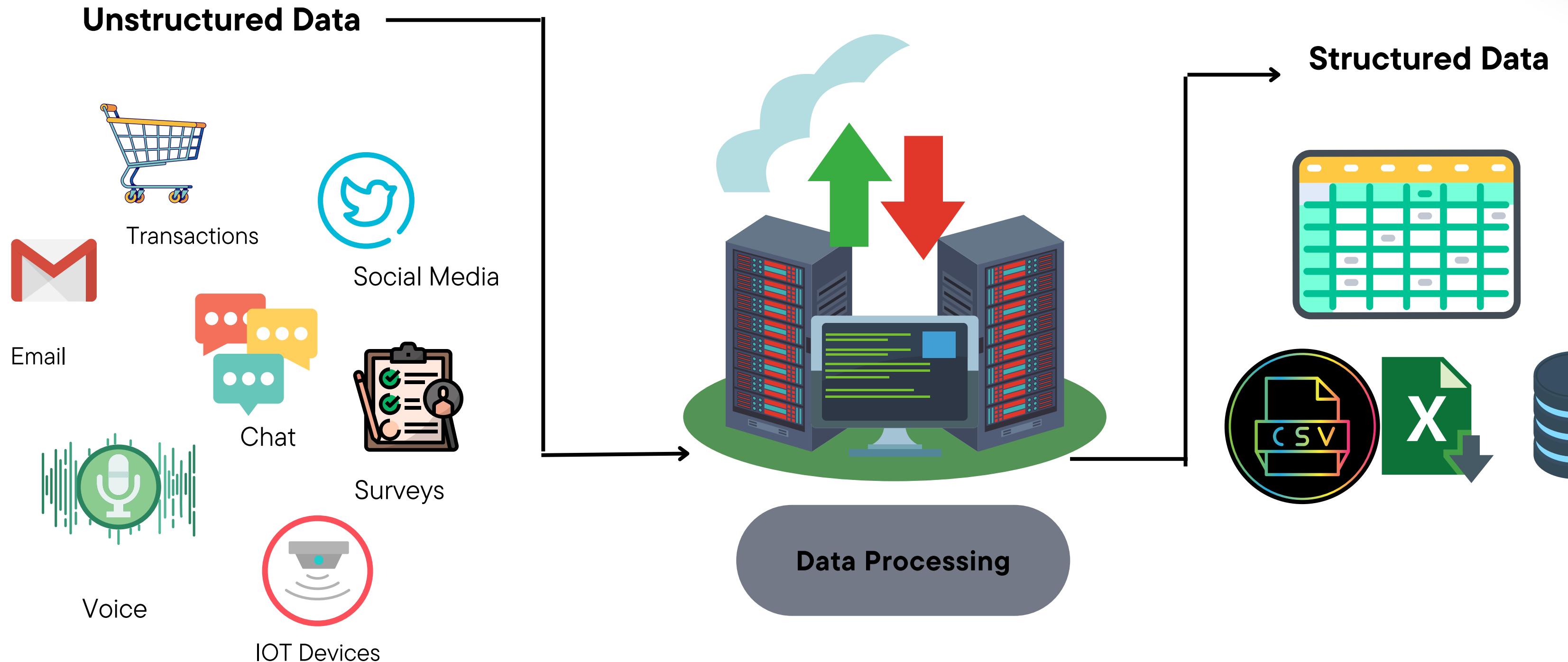


1

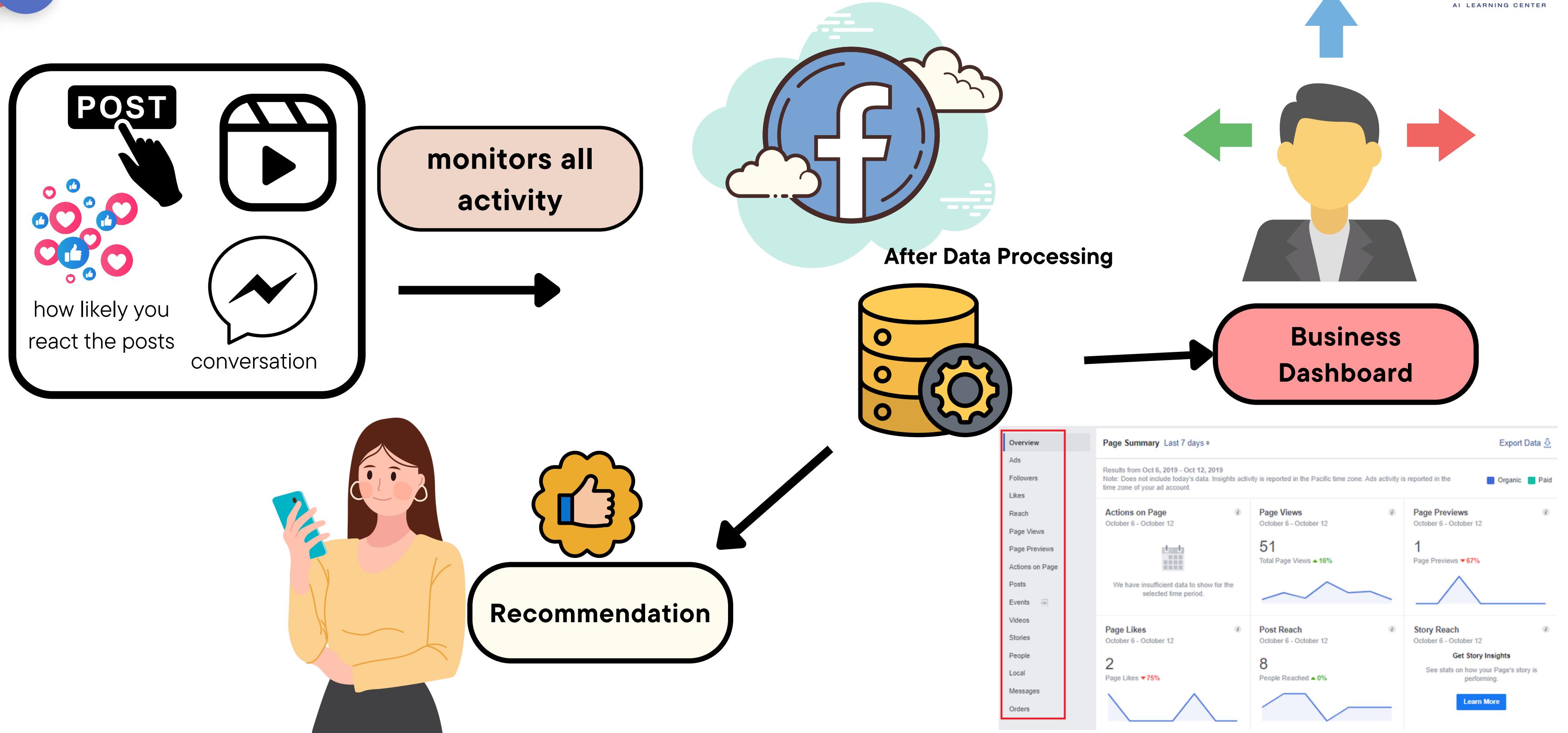
DATA vs INFORMATION vs INSIGHTS



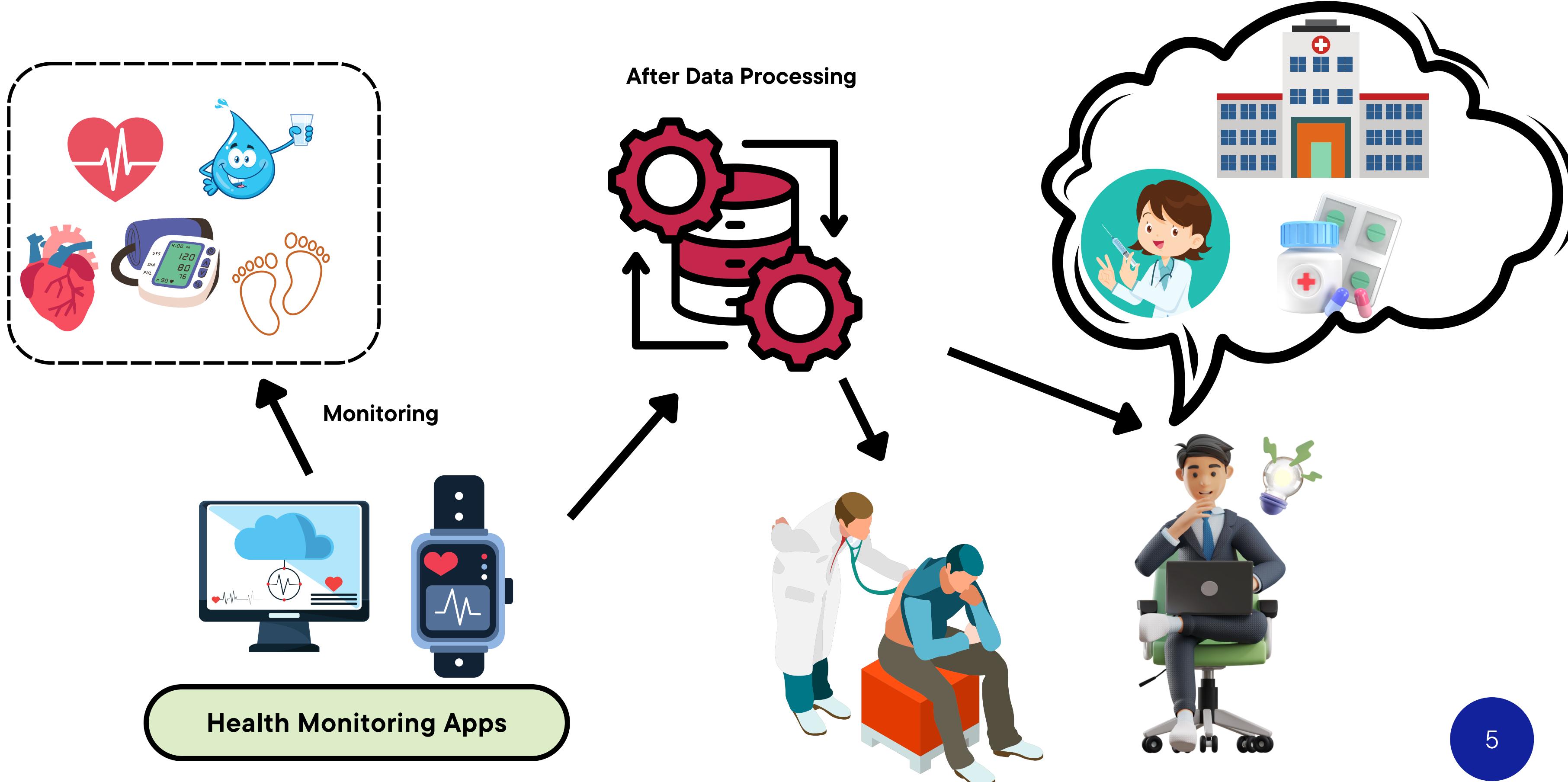
DATA TYPES



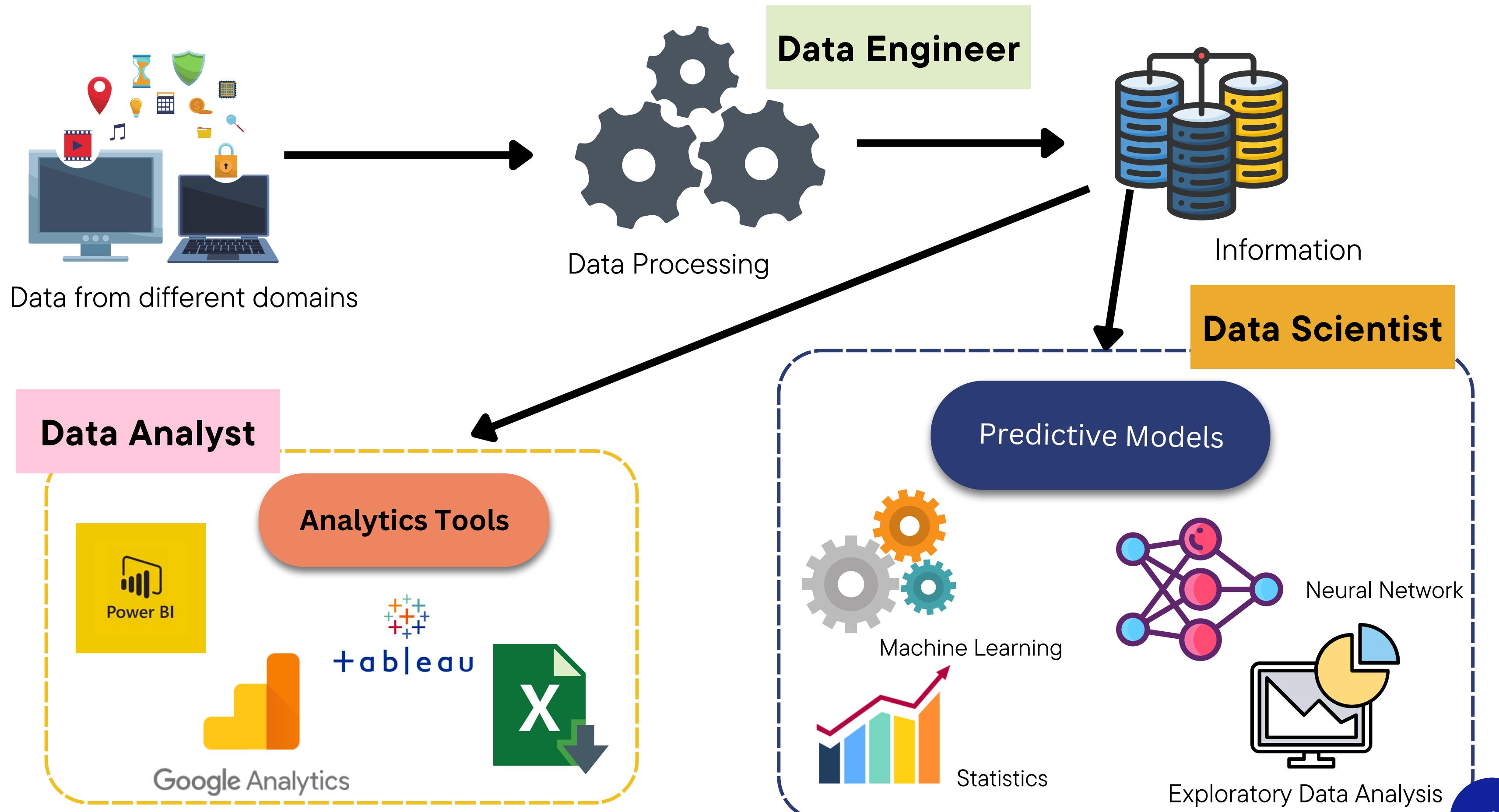
Case Study 1# Facebook Data Analytics



Case Study #2 : Future Trend in Healthcare



Then, How to Extract Insights ?



2

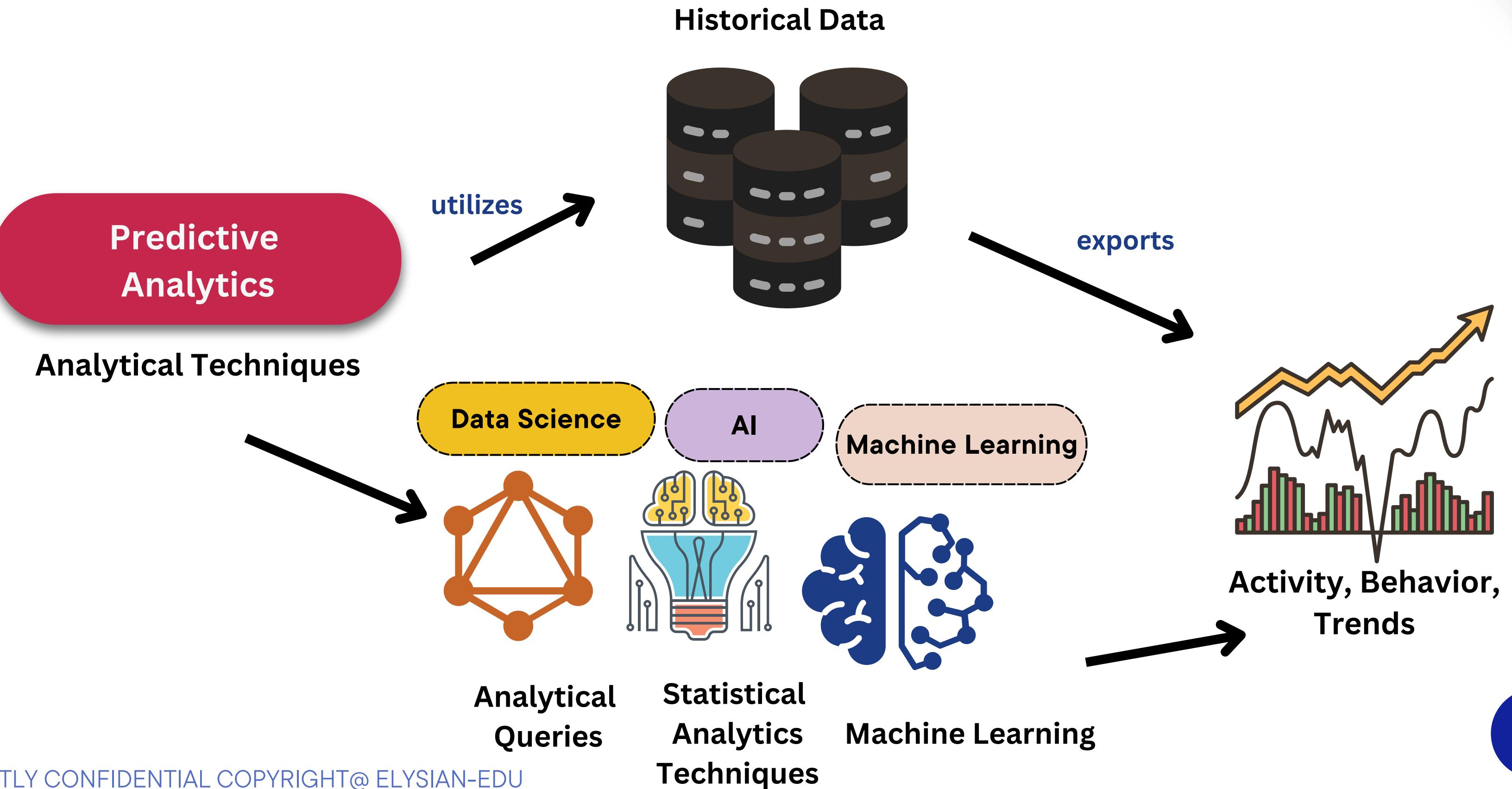


How enterprises are using Predictive Analytics to transform historical data into future insights

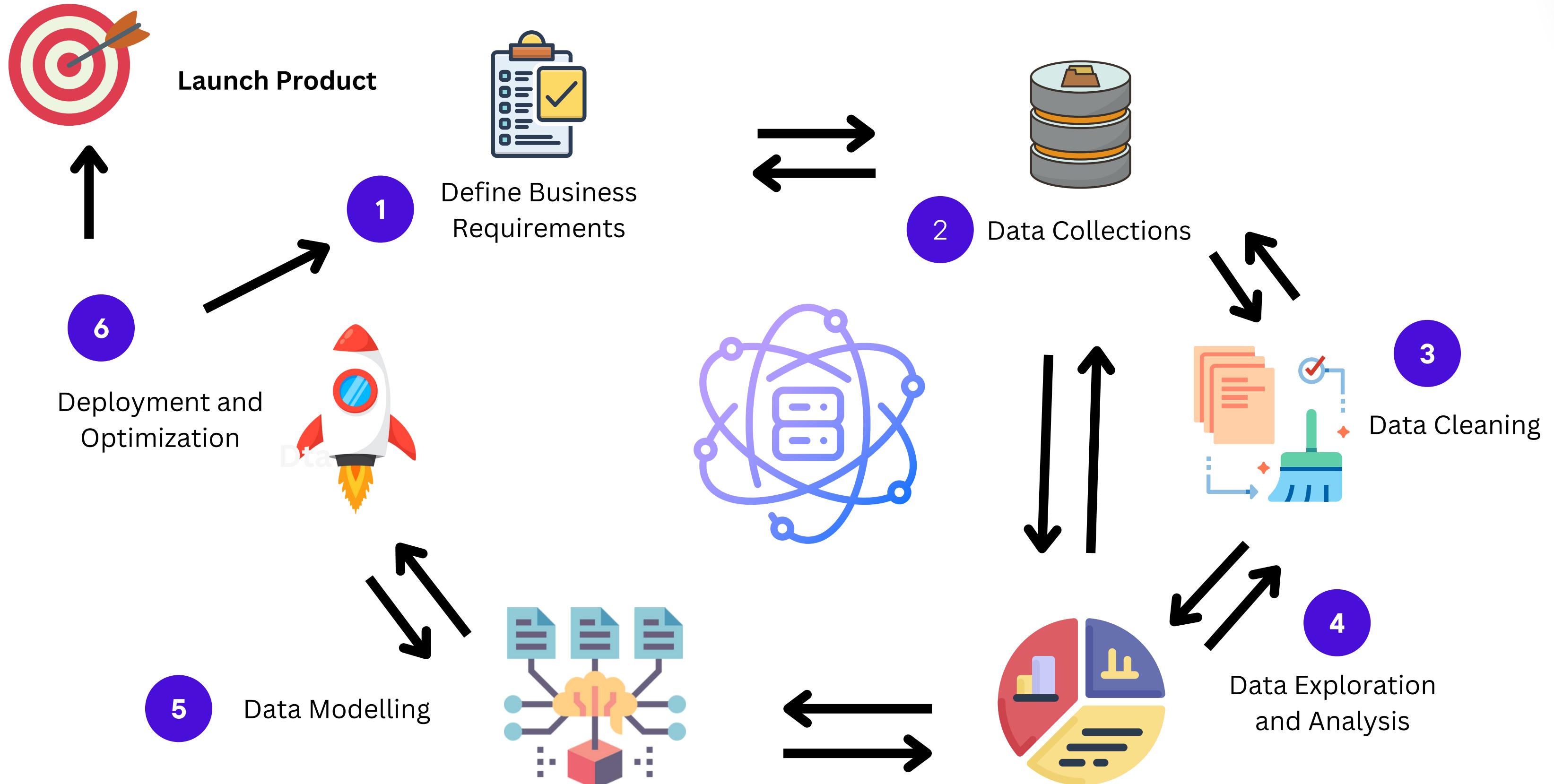
Here we go.....



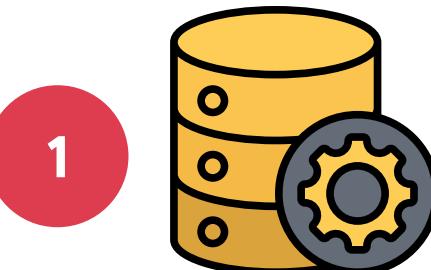
What is Predictive Analytics?



Data Science project LIFE CYCLE



Challenges



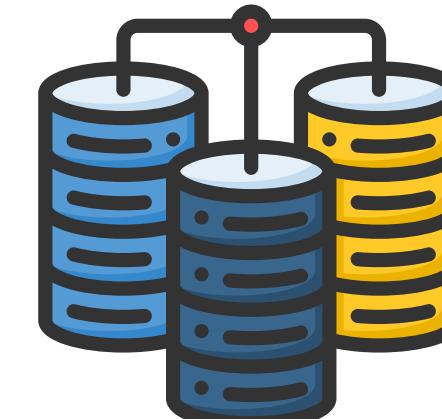
1 Data Collection

- organizations collect huge amounts of data without doing anything to determine whether it's useful or not.
- abundance of data sources, which makes it difficult to find the right data



2 Data Preparation

- spend a lot of time on processing and preparing data
- the worst part of any ML project, it is a crucial process that ensures ML models are built on high-quality data



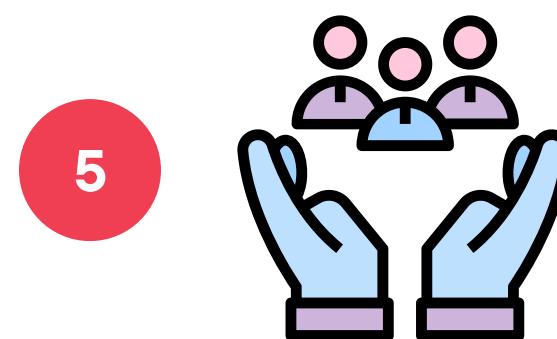
3 Managing Large Data Volume

- most of this data is unstructured
- organizations are more frequently turning to big data platforms to process for storage, management, cleansing, and analytics so that they can extract the insights



4 Extracting Right insights

- Organizations want to use their data to achieve their goals, and the only way to do this is by extracting relevant insights from it so that leaders can use it to make their decisions.



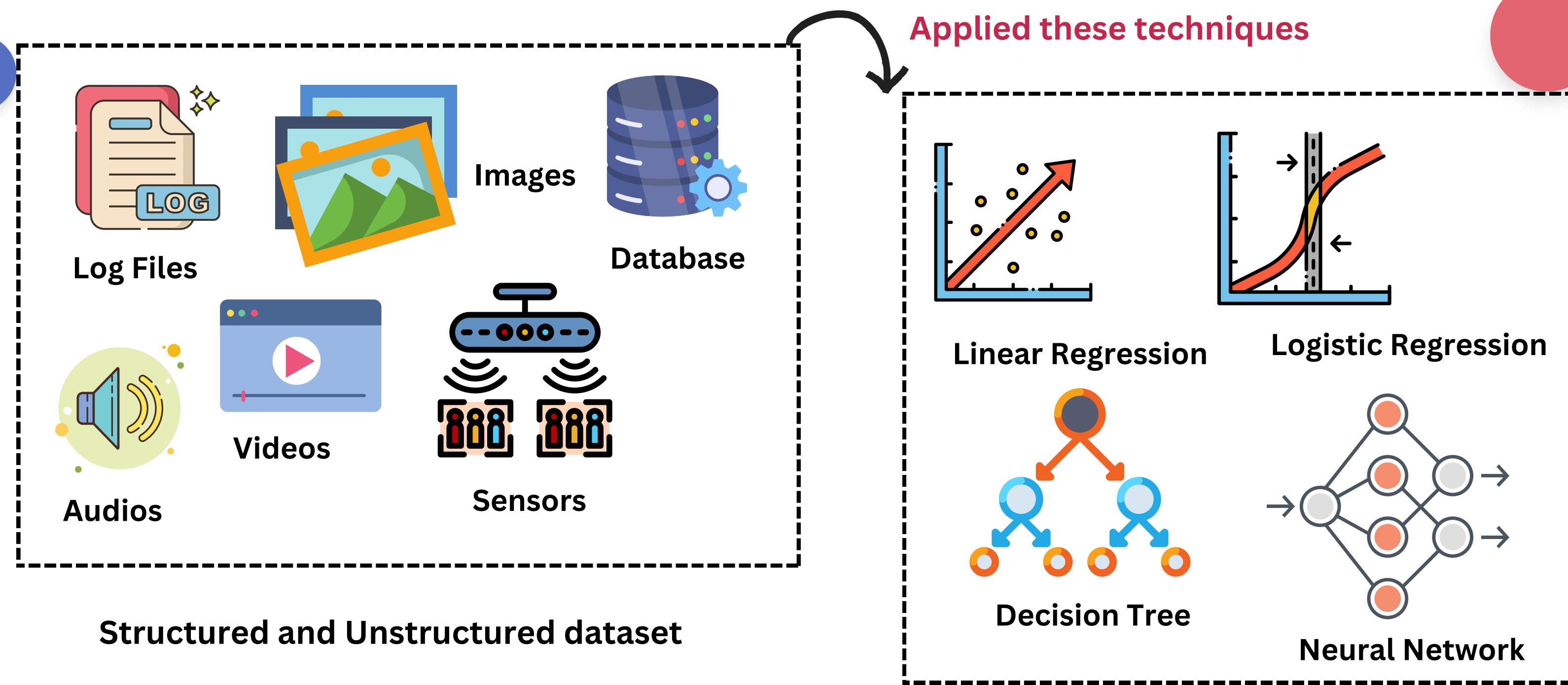
5 Lack of Human Resource

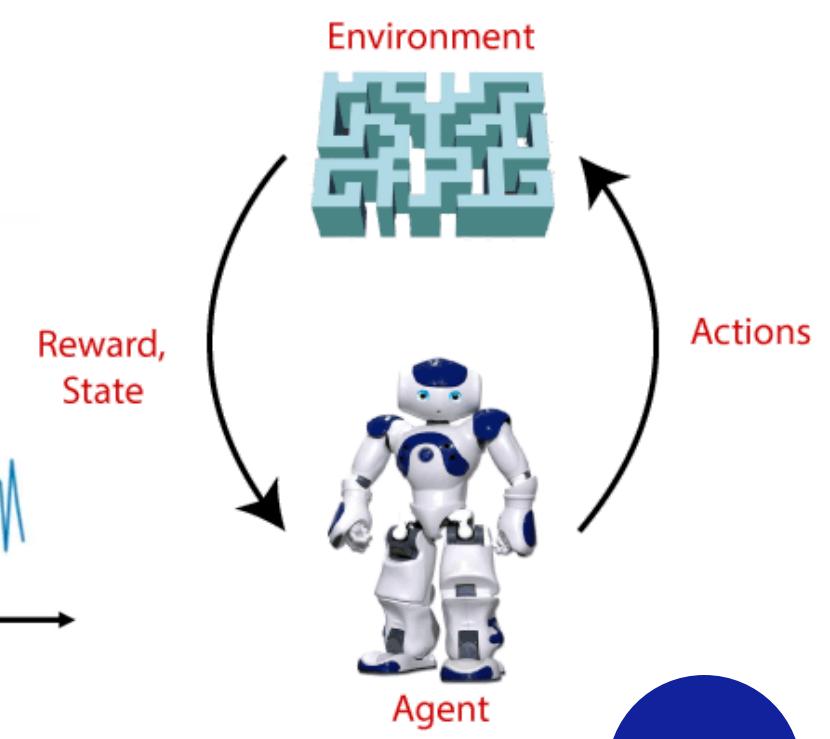
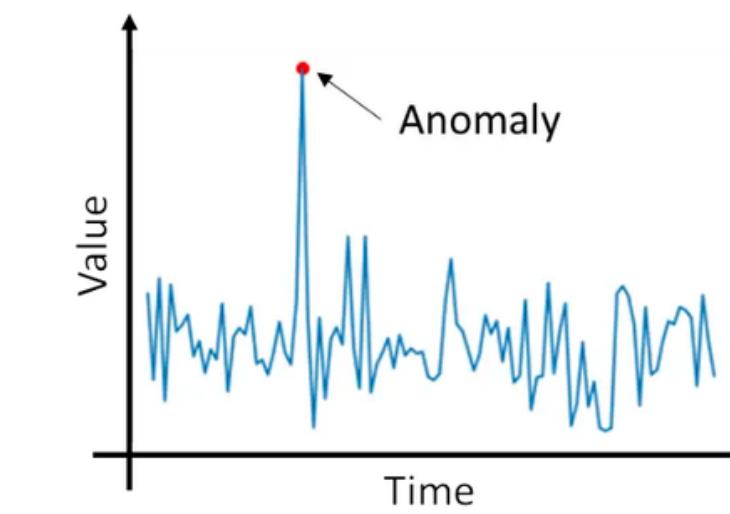
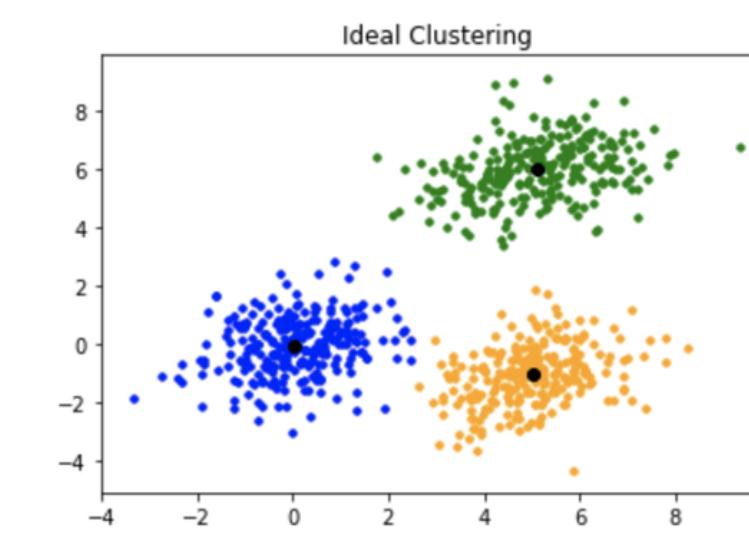
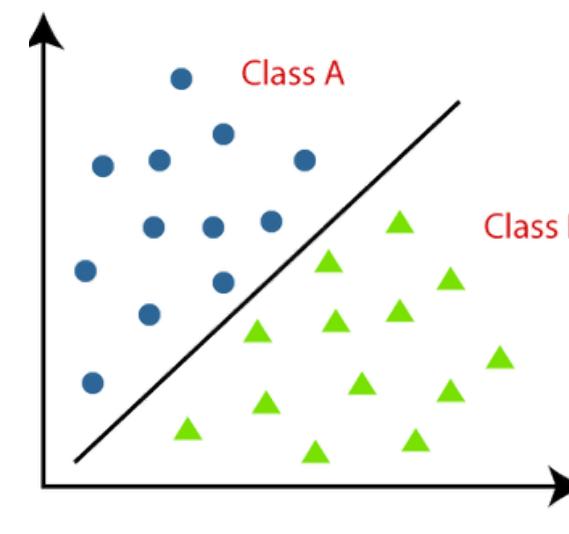
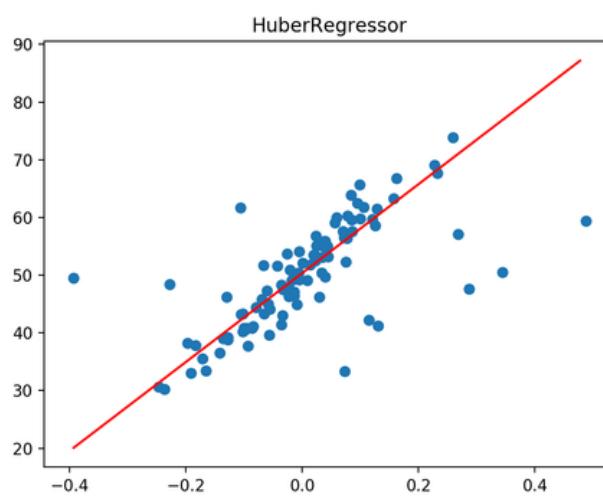
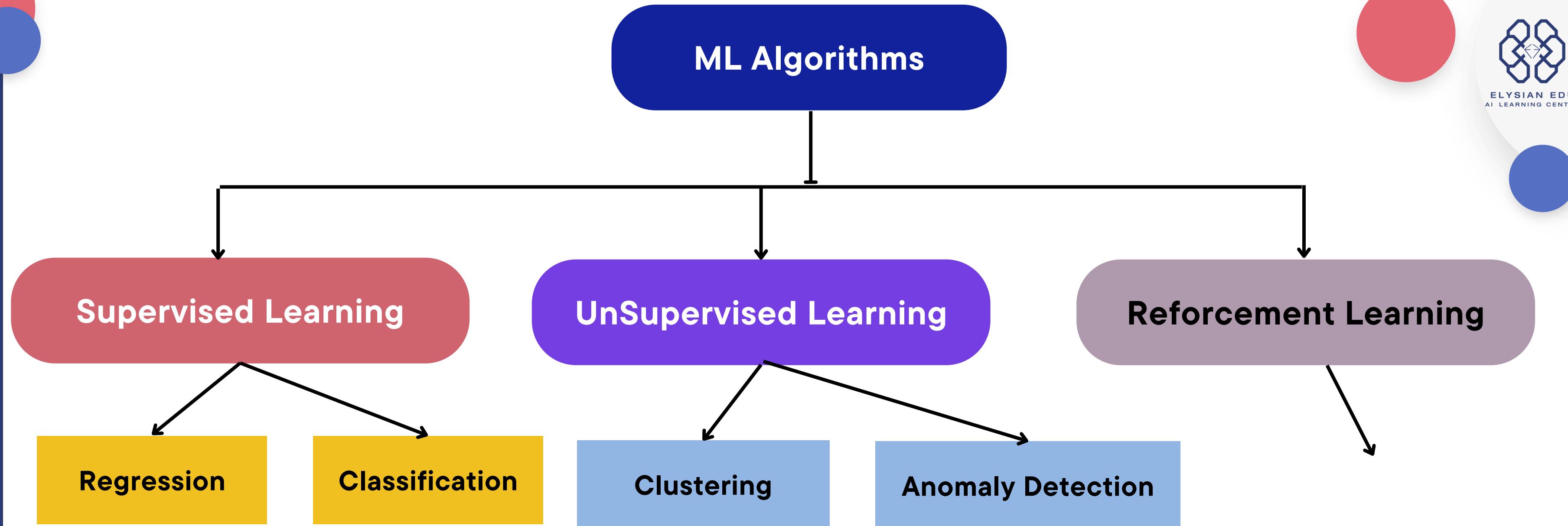
- Organizations often struggle to find the right people with the right level of knowledge and domain expertise to put together their ML teams
- also struggle to find people who have the right business perspective and decision using data science



6 Communication

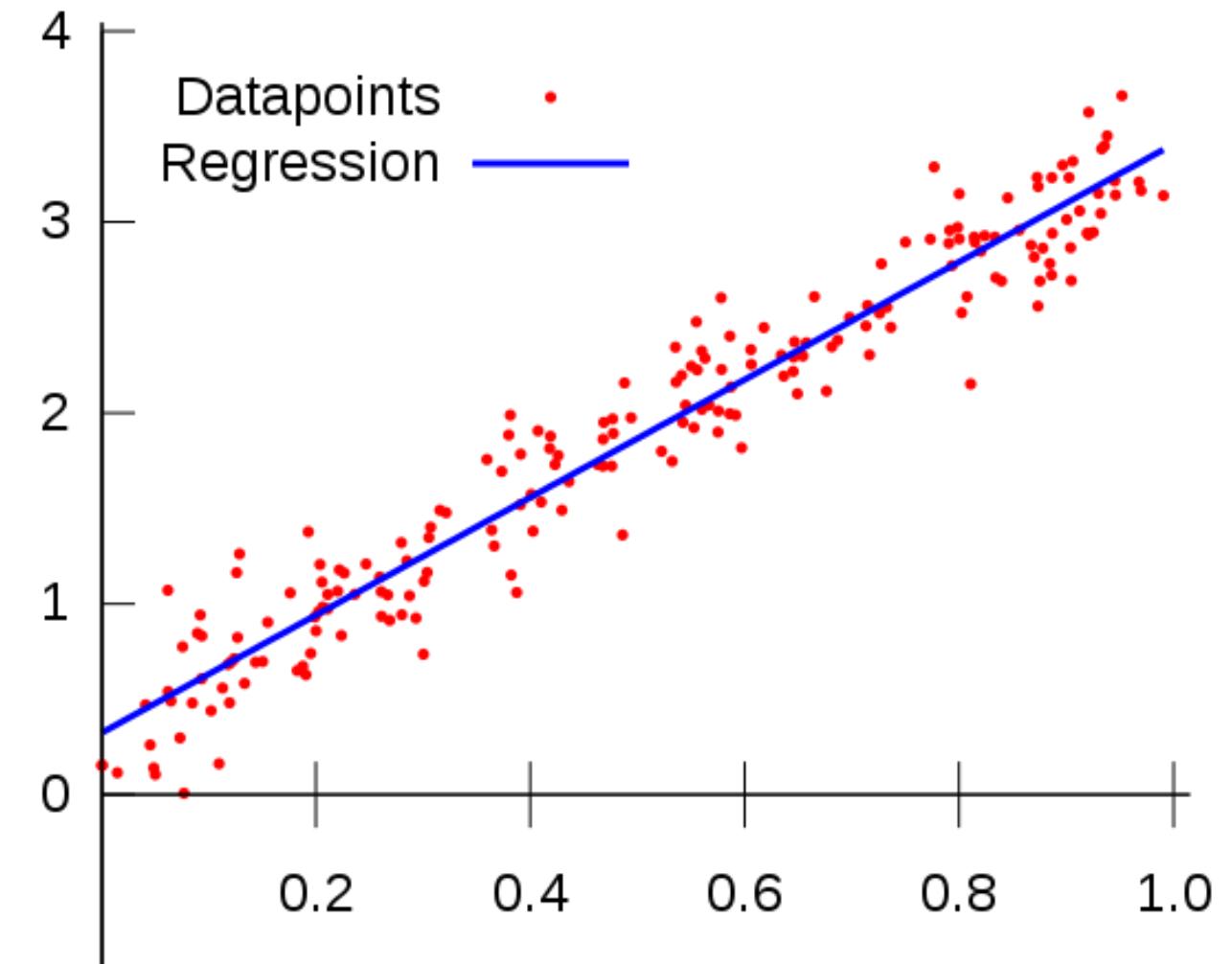
- It's common to work alongside other analysts or scientists as part of a team, especially when working on large projects or handling big datasets.
- It's important to be a good communicator in order to work effectively with others.





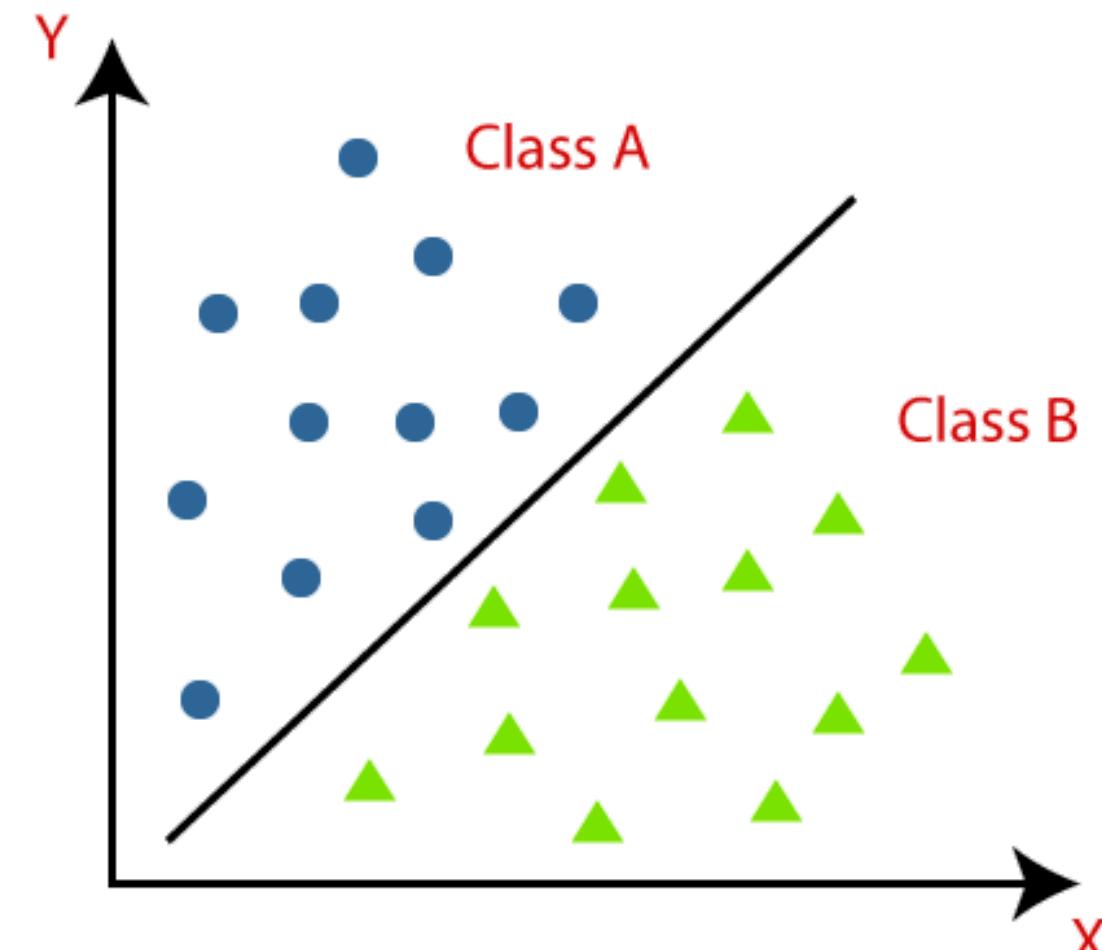
Supervised : Regression Models

- Supervised Method
- Regression is a technique for investigating the relationship between independent variables or features and a dependent variable or outcome.
- mainly used in predictive analytics to forecast trends and predict outcomes



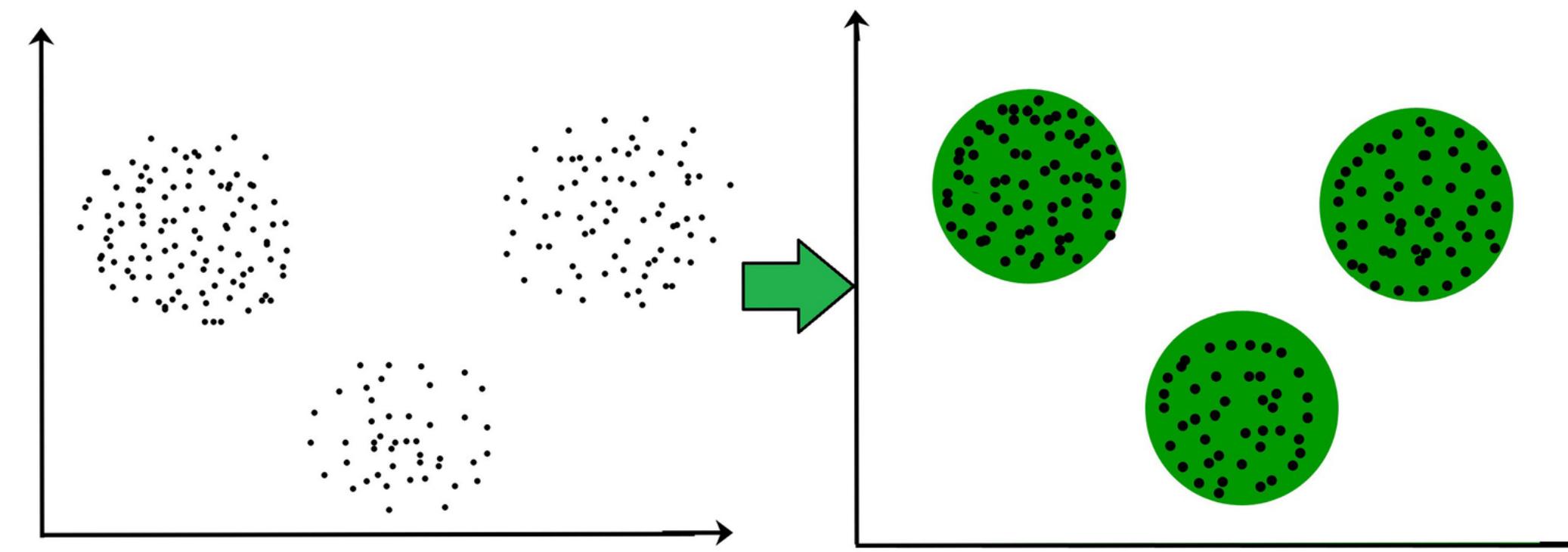
Supervised : Classification

- Supervised Learning technique that is used to **identify the category of new observations** on the basis of training data.
- A model **will use the training dataset** and will calculate how to best map examples of input data to specific class labels.
- **the training dataset must be sufficiently representative of the problem** and have many examples of each class label.



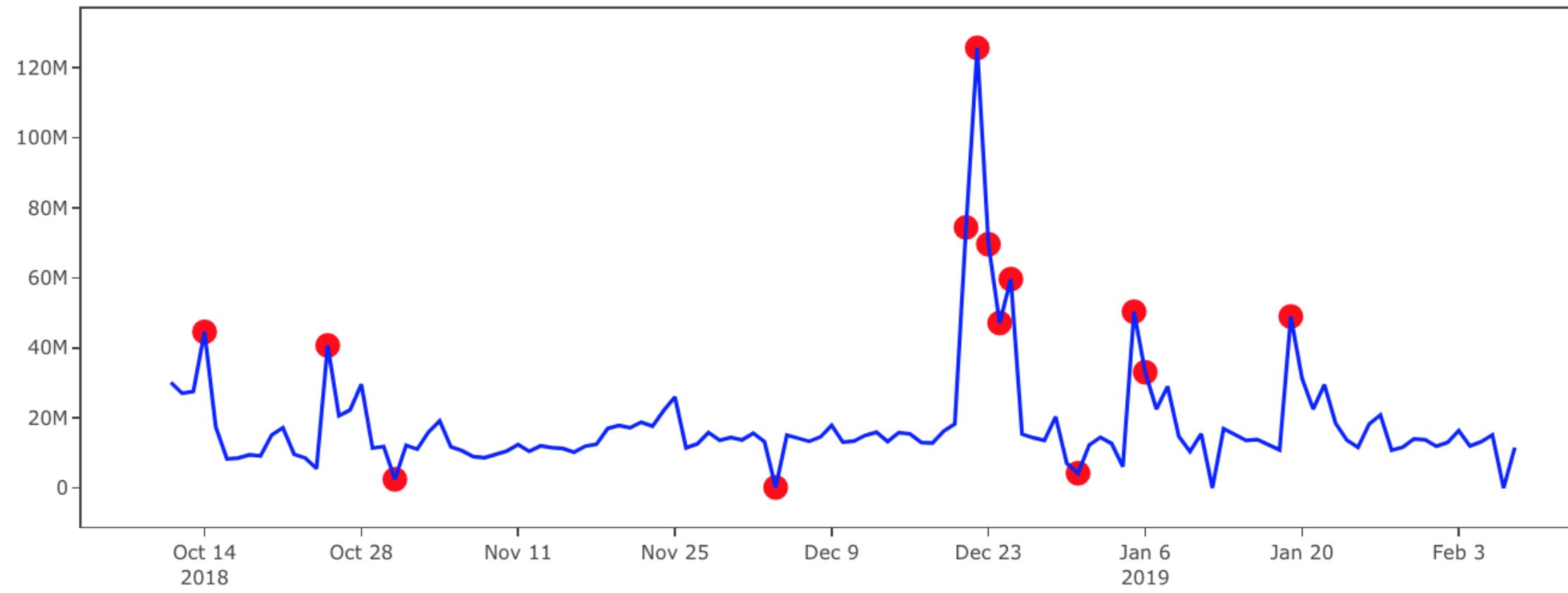
Unsupervised : Clustering

- An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labeled responses.
- Clustering is the task of dividing the population or data points into a number of groups into similar and dissimilar groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.



Unsupervised : Anomaly Detection

- Unsupervised Method
- The process of identifying unexpected items or events in data sets, which differ from the norm.
- Anomaly detection has two basic assumptions:
 - Anomalies only occur very rarely in the data.
 - Their features differ from the normal instances significantly.



Business Use Cases

Regression

- Financial forecasting (like house price estimates, or stock prices)
- Sales and promotions forecasting
- Weather analysis and prediction
- Time series forecasting

Classification

- Sentiment Analysis
- Malware Classification
- Spam Classification

Clustering

- Marketing
- Recommender System
- Libraries
- Insurance

Anomaly Detection

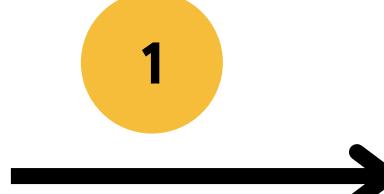
- Network intrusion detection systems
- System Health Monitoring
- Event detection in sensor networks
- Defect detection in images using machine vision



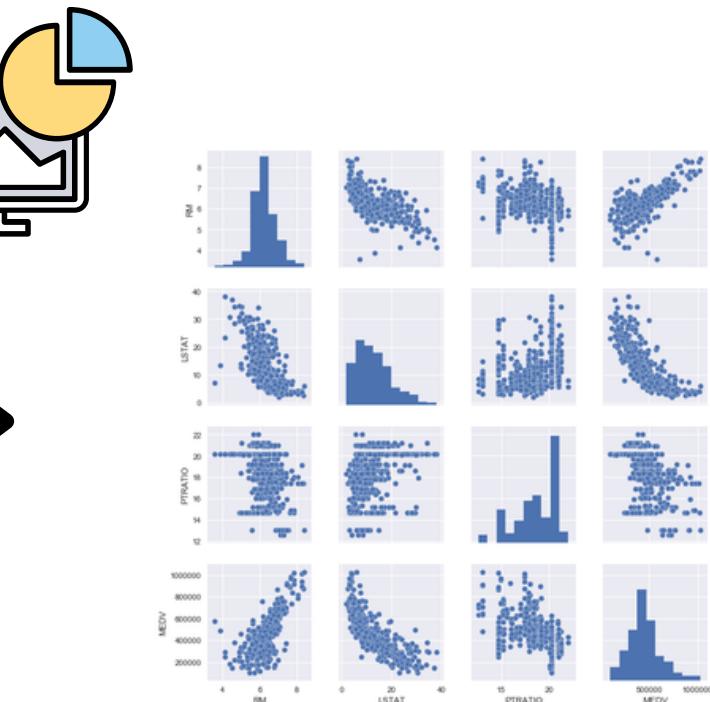
Case Study # Housing Price using Regression

| | CRIM | ZN | INDUS | CHAS | NOX | RM | AGE | DIS | RAD | TAX | PTRATIO | B | LSTAT |
|---|---------|------|-------|------|-------|-------|-------|--------|-----|-------|---------|--------|-------|
| 0 | 0.00632 | 18.0 | 2.31 | 0.0 | 0.538 | 6.575 | 65.2 | 4.0900 | 1.0 | 296.0 | 15.3 | 396.90 | 4.98 |
| 1 | 0.02731 | 0.0 | 7.07 | 0.0 | 0.469 | 6.421 | 78.9 | 4.9671 | 2.0 | 242.0 | 17.8 | 396.90 | 9.14 |
| 2 | 0.02729 | 0.0 | 7.07 | 0.0 | 0.469 | 7.185 | 61.1 | 4.9671 | 2.0 | 242.0 | 17.8 | 392.83 | 4.03 |
| 3 | 0.03237 | 0.0 | 2.18 | 0.0 | 0.458 | 6.998 | 45.8 | 6.0622 | 3.0 | 222.0 | 18.7 | 394.63 | 2.94 |
| 4 | 0.06905 | 0.0 | 2.18 | 0.0 | 0.458 | 7.147 | 54.2 | 6.0622 | 3.0 | 222.0 | 18.7 | 396.90 | 5.33 |
| 5 | 0.02985 | 0.0 | 2.18 | 0.0 | 0.458 | 6.430 | 58.7 | 6.0622 | 3.0 | 222.0 | 18.7 | 394.12 | 5.21 |
| 6 | 0.08829 | 12.5 | 7.87 | 0.0 | 0.524 | 6.012 | 66.6 | 5.5605 | 5.0 | 311.0 | 15.2 | 395.60 | 12.43 |
| 7 | 0.14455 | 12.5 | 7.87 | 0.0 | 0.524 | 6.172 | 96.1 | 5.9505 | 5.0 | 311.0 | 15.2 | 396.90 | 19.15 |
| 8 | 0.21124 | 12.5 | 7.87 | 0.0 | 0.524 | 5.631 | 100.0 | 6.0821 | 5.0 | 311.0 | 15.2 | 386.63 | 29.93 |
| 9 | 0.17004 | 12.5 | 7.87 | 0.0 | 0.524 | 6.004 | 85.9 | 6.5921 | 5.0 | 311.0 | 15.2 | 386.71 | 17.10 |

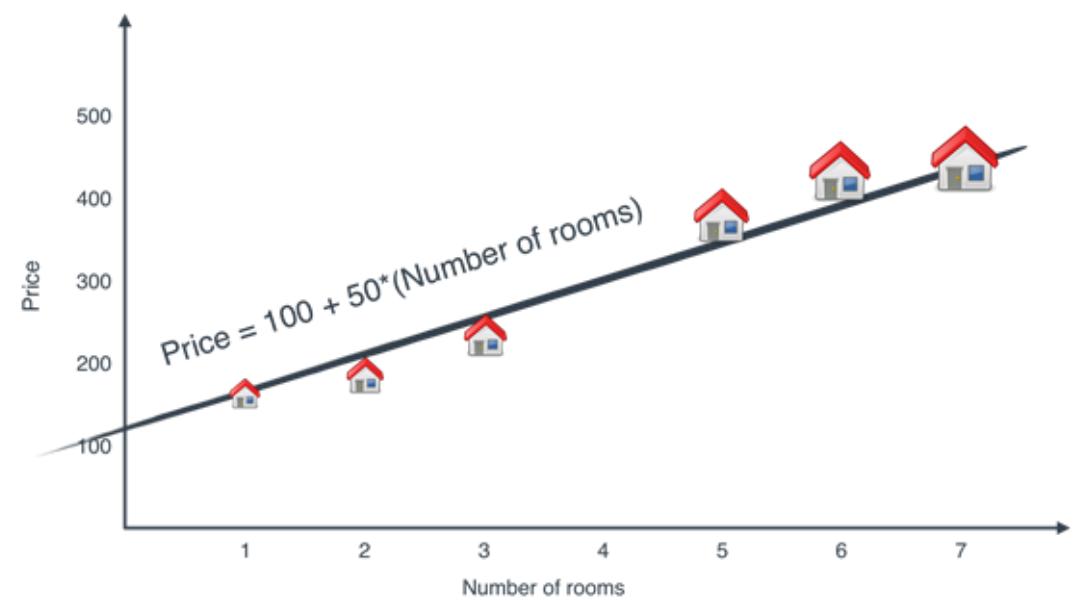
Dataset



Data Preprocessing



Exploratory Data Analysis



Output



Model Training

3

Data Mining Using Graph Representation

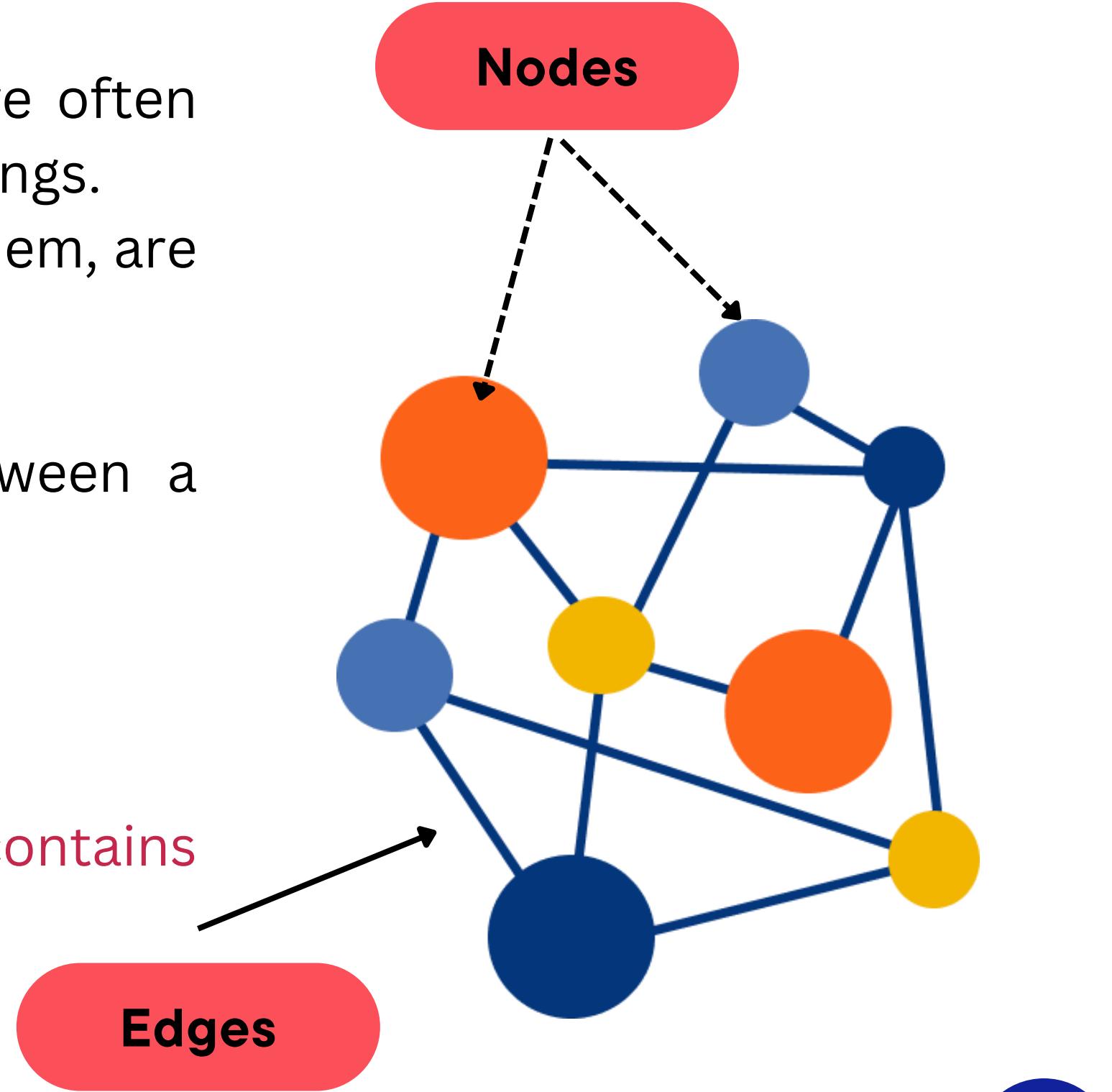
Here we go.....

A Machine Learning Innovation in Predictive Analytics



What is Graph Representation ?

- Graphs are all around us; real world objects are often defined in terms of their connections to other things.
- A set of objects, and the connections between them, are naturally expressed as a graph
- A graph represents the relations (edges) between a collection of entities (nodes).
 - Vertex (or node) attributes
 - Edge (or link) attributes and directions
 -
- primary goal is to learn an embedding that contains information about its neighborhood.



How data transform into Graph

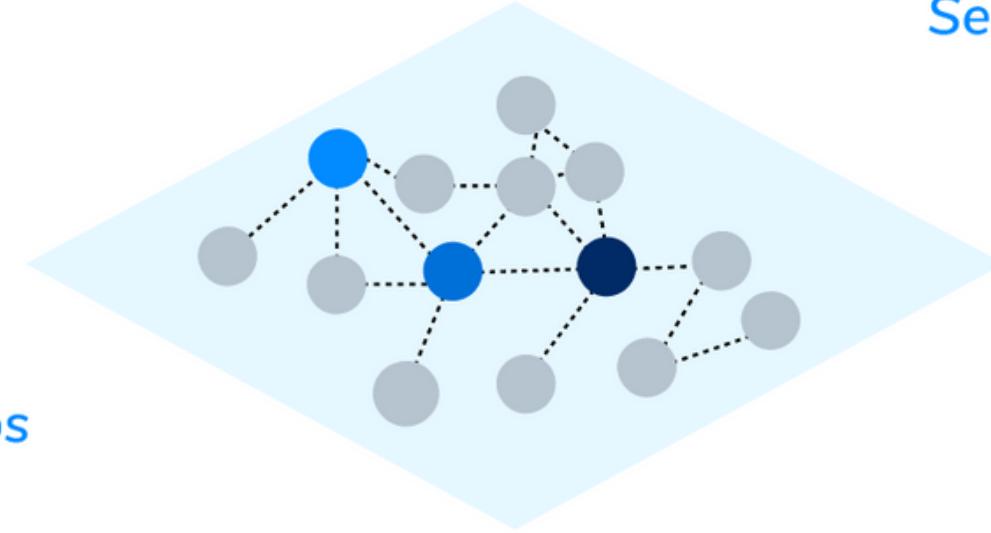


Data

Bridge together **diverse** and **disparate data** regardless of data type, such as structured, unstructured, and semi-structured.



Relationships



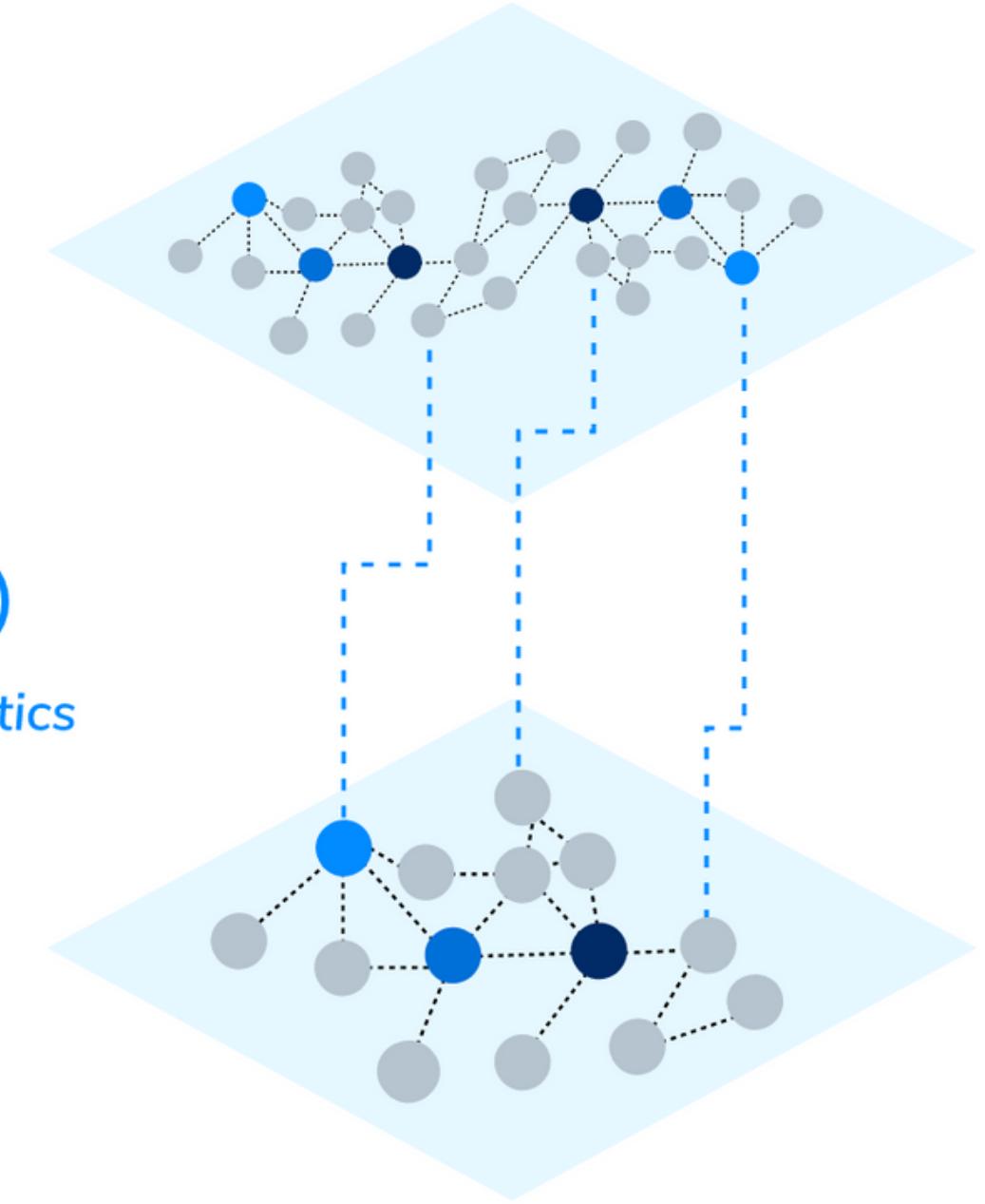
Graph

Dynamic Context

Map data and draw connections among them for the first layer of dynamic context, which provides immediate understanding.



Semantics

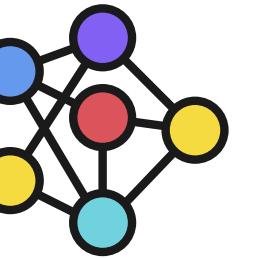


Knowledge Graph

Deep Dynamic Context

Apply semantics to provide deeper context to connected data. The deeper the context, the more powerful the insights.

3.1



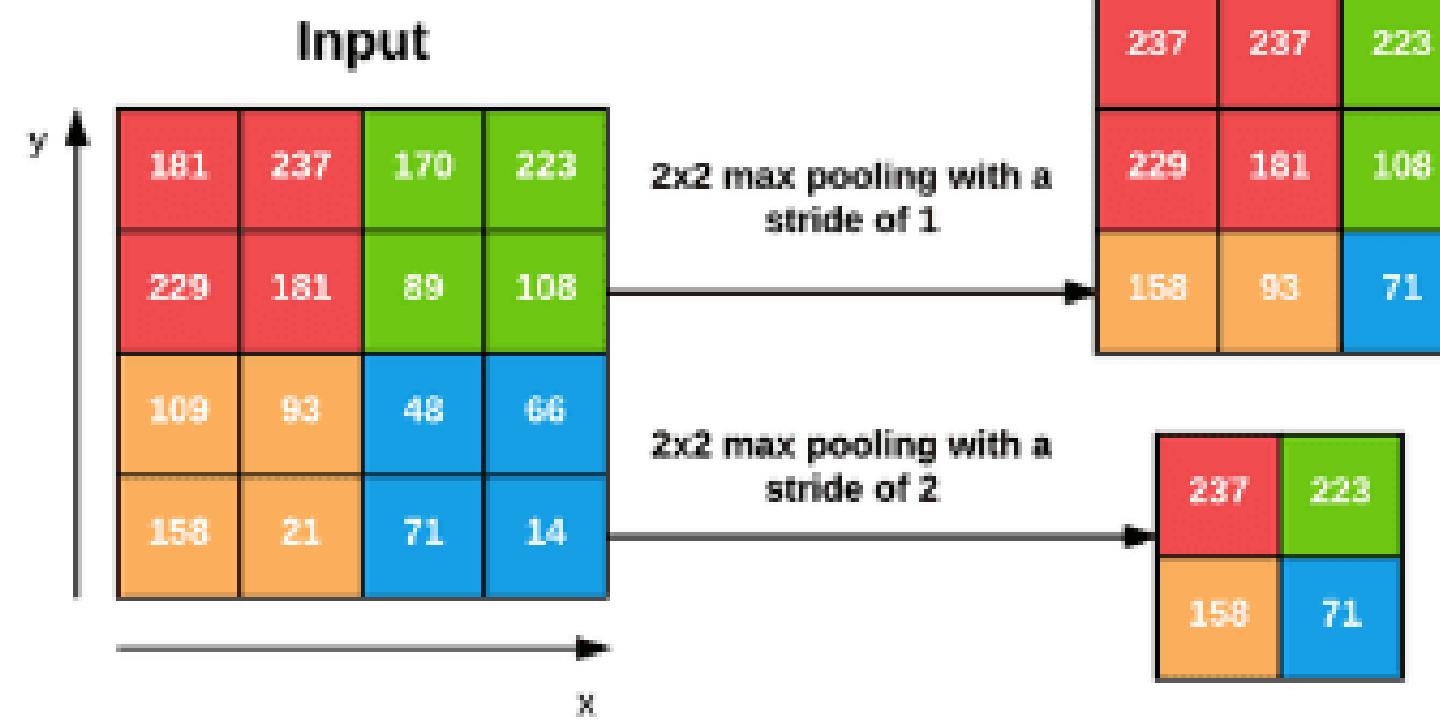
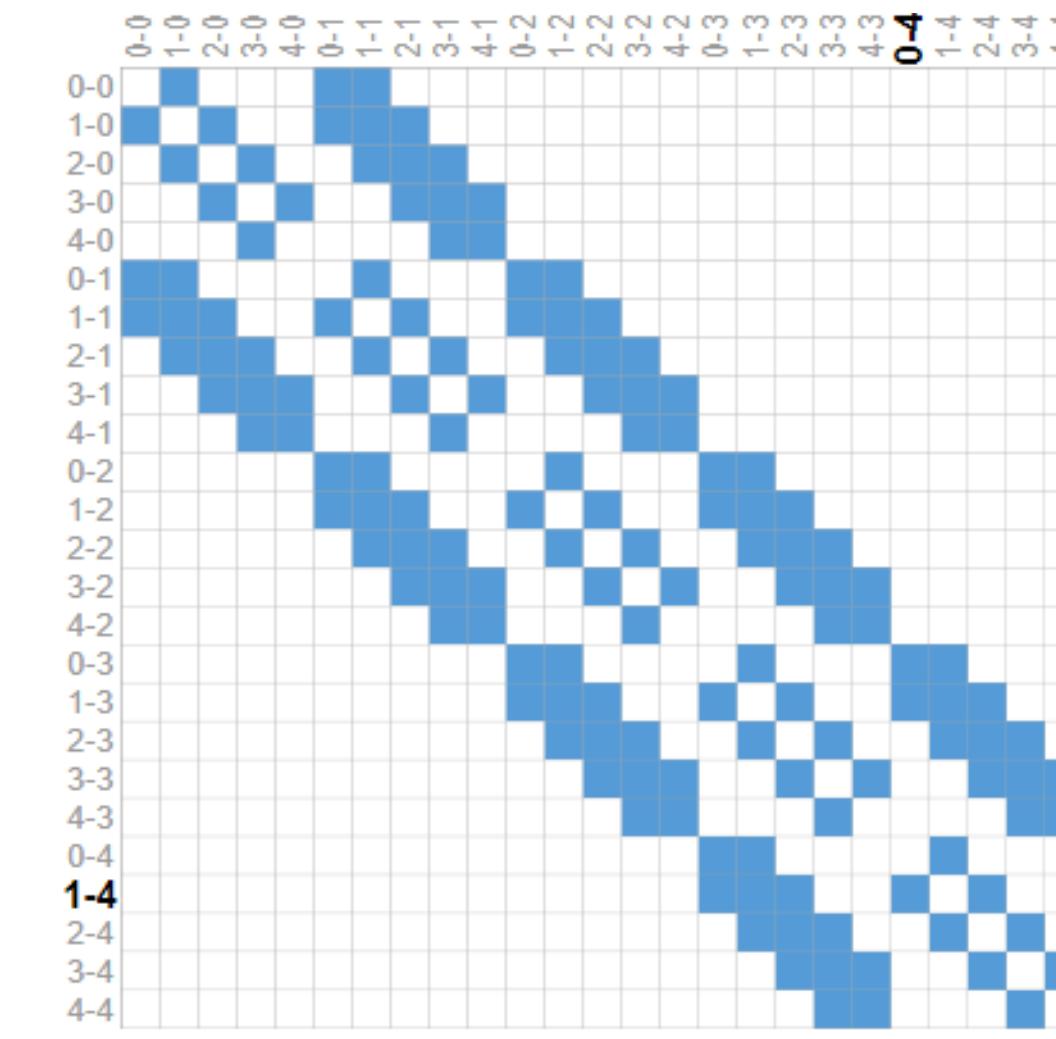
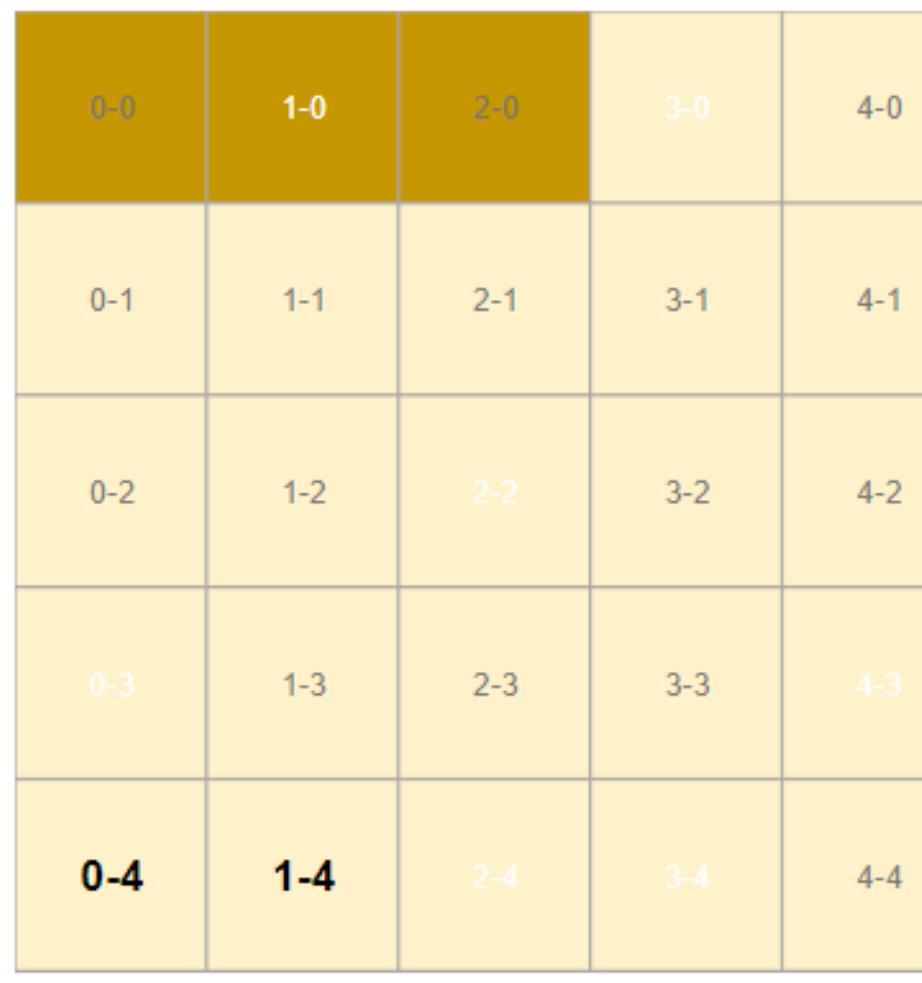
From Image to Graph

CNN to GNN

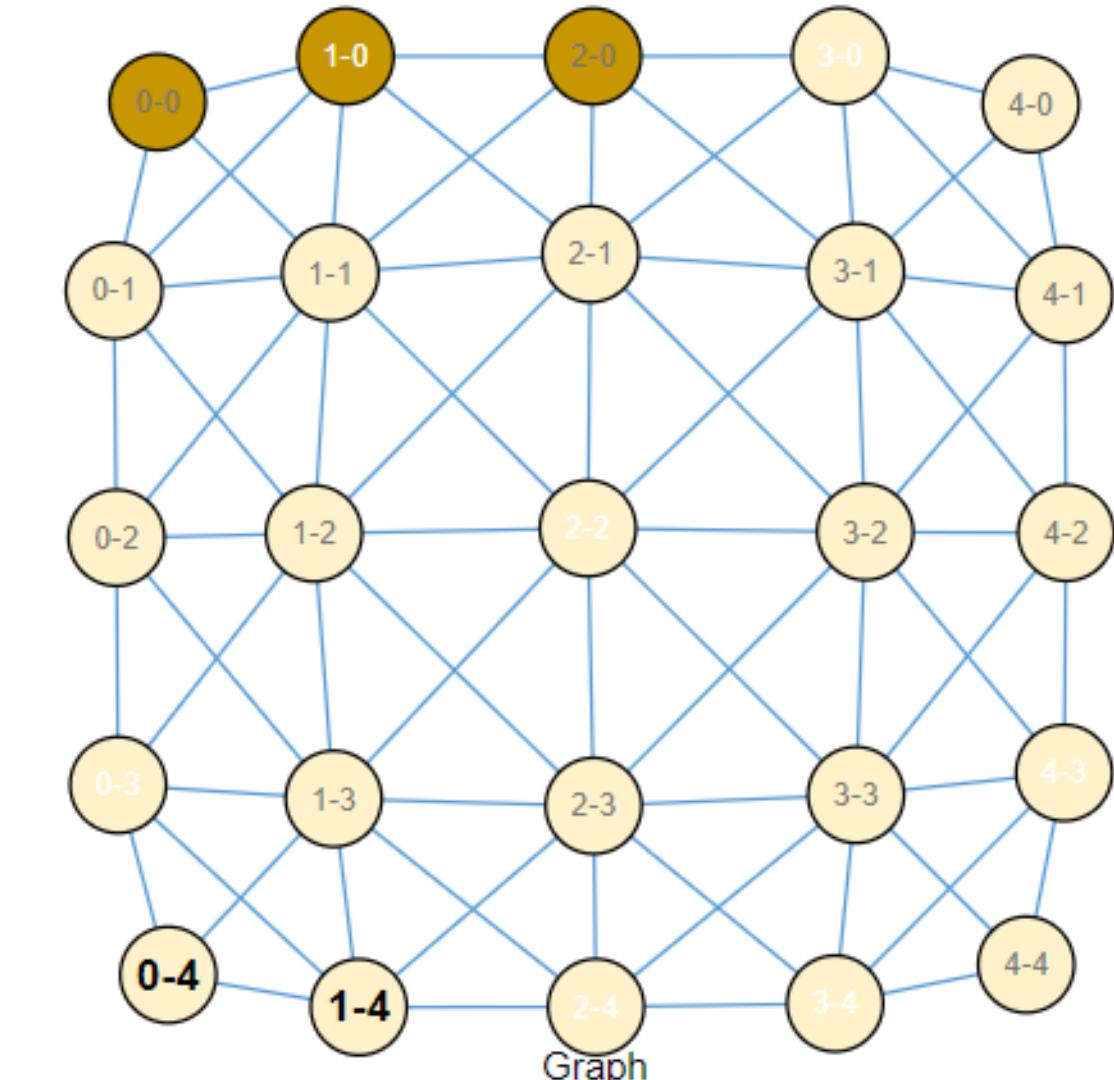
Here we go.....

A Machine Learning Innovation in Predictive Analytics

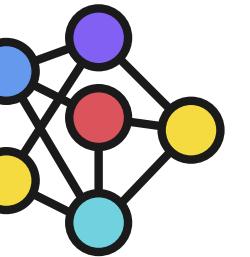



CNN

GNN


Adjacency Matrix



3.2



From Text to Knowledge

Here we go.....

A Machine Learning Innovation in Predictive Analytics



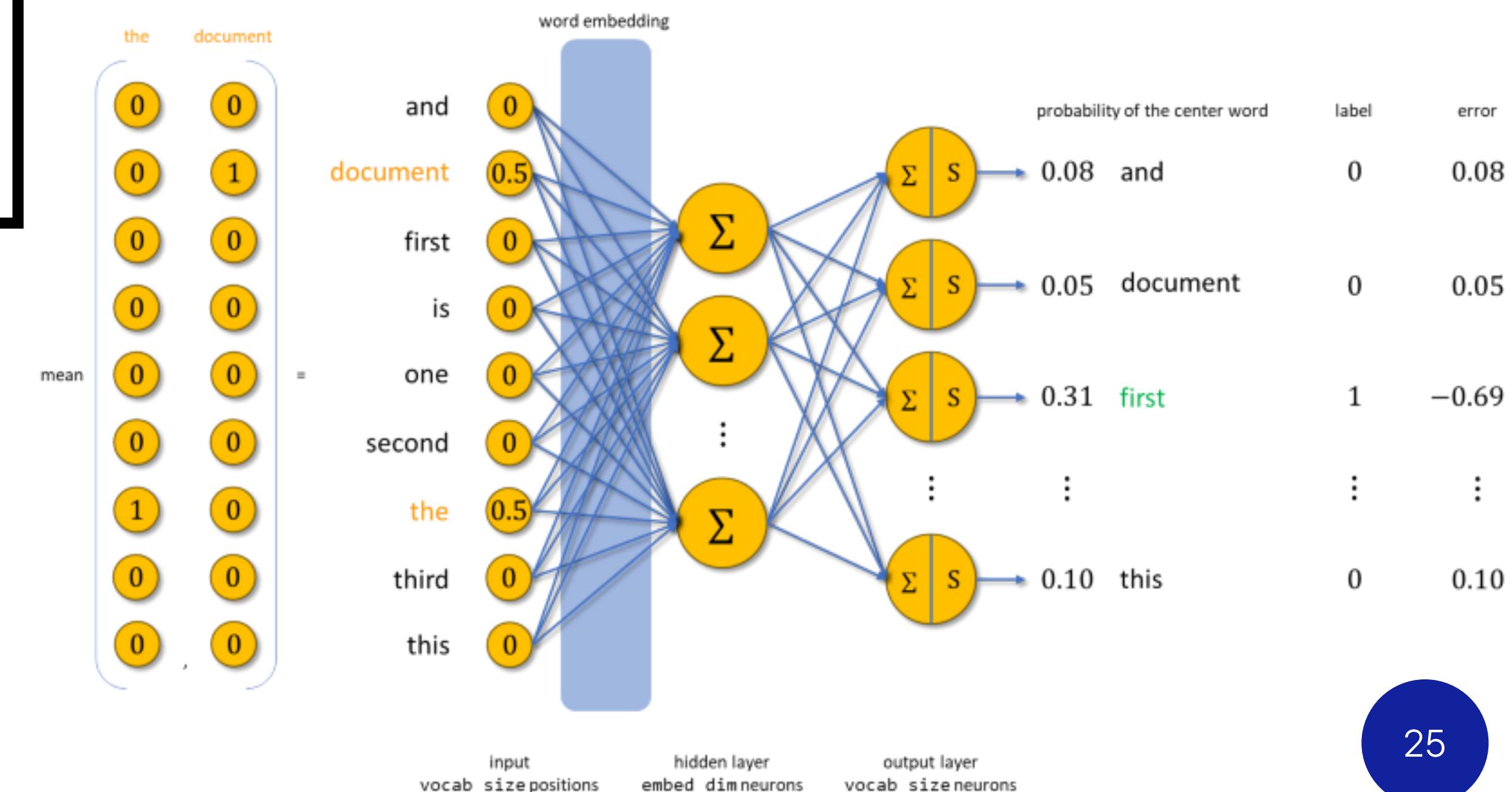
with Bags-of-Words

Corpus

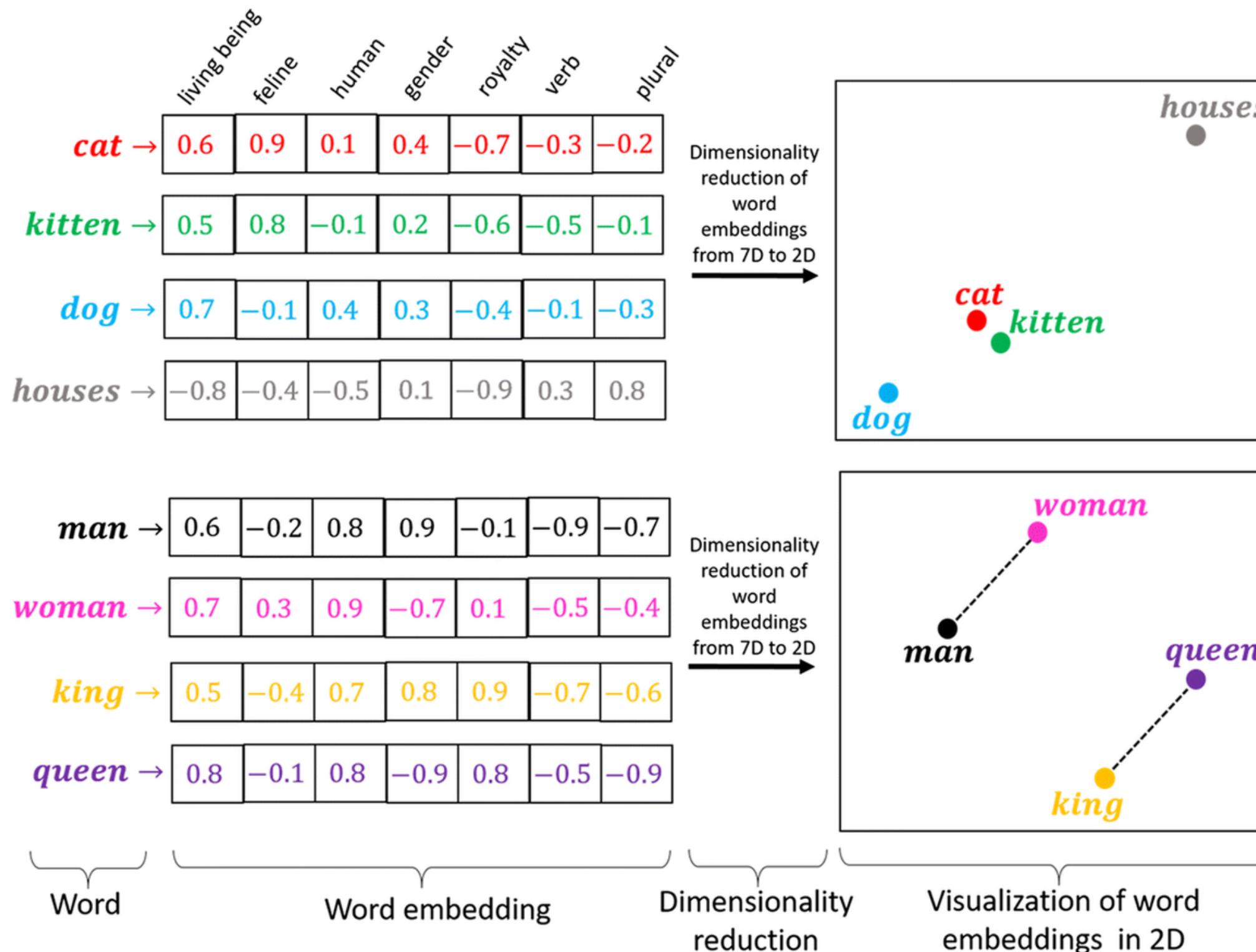
| and | document | first | is | one | second | the | third | this |
|-----|----------|-------|----|-----|--------|-----|-------|------|
| 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| 0 | 1 | 0 | 1 | 0 | 2 | 1 | 0 | 1 |
| 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |

1. This is the first document.
2. This is the second document.
3. And the third one.
4. Is this the first document?

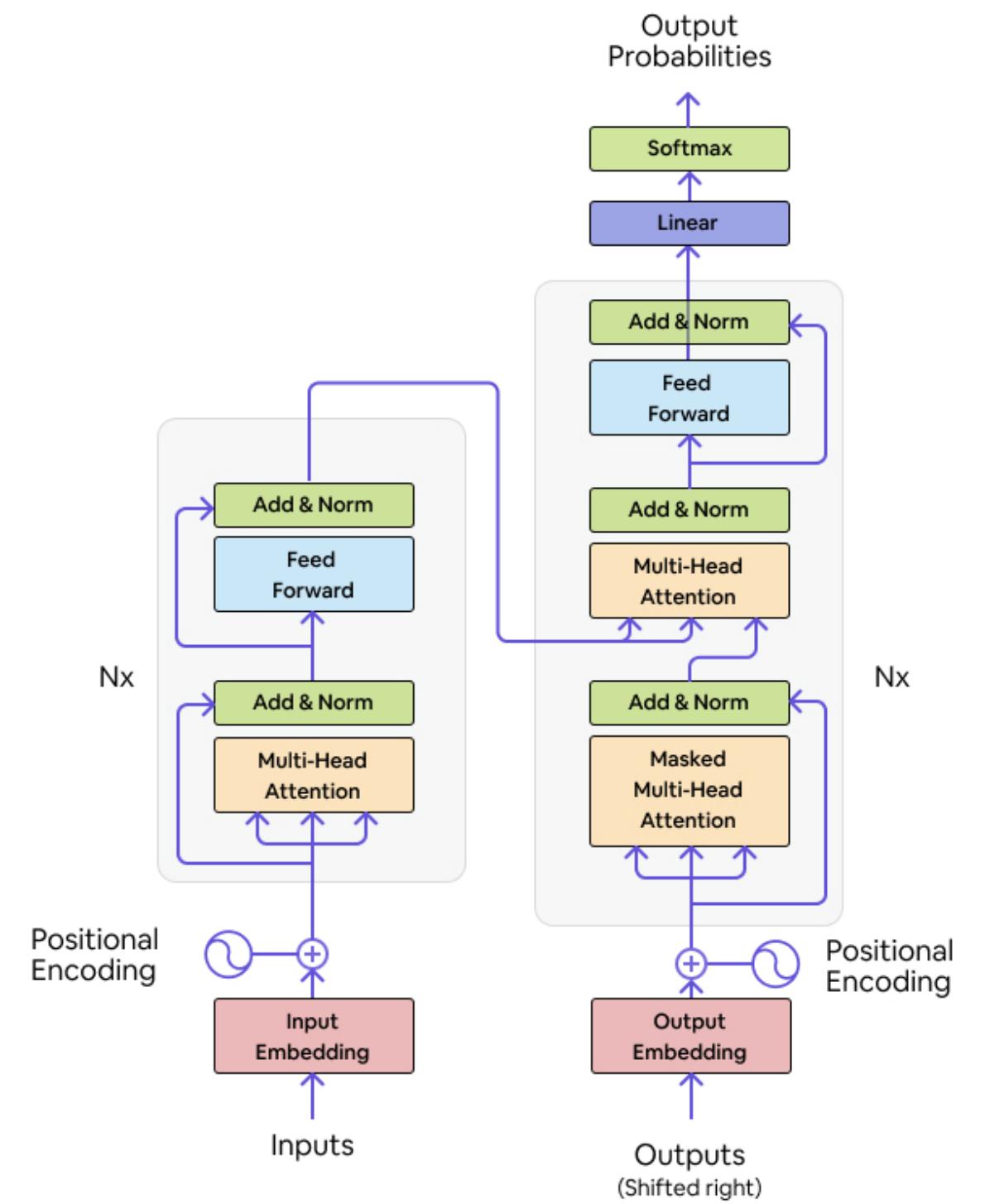
Word2Vec



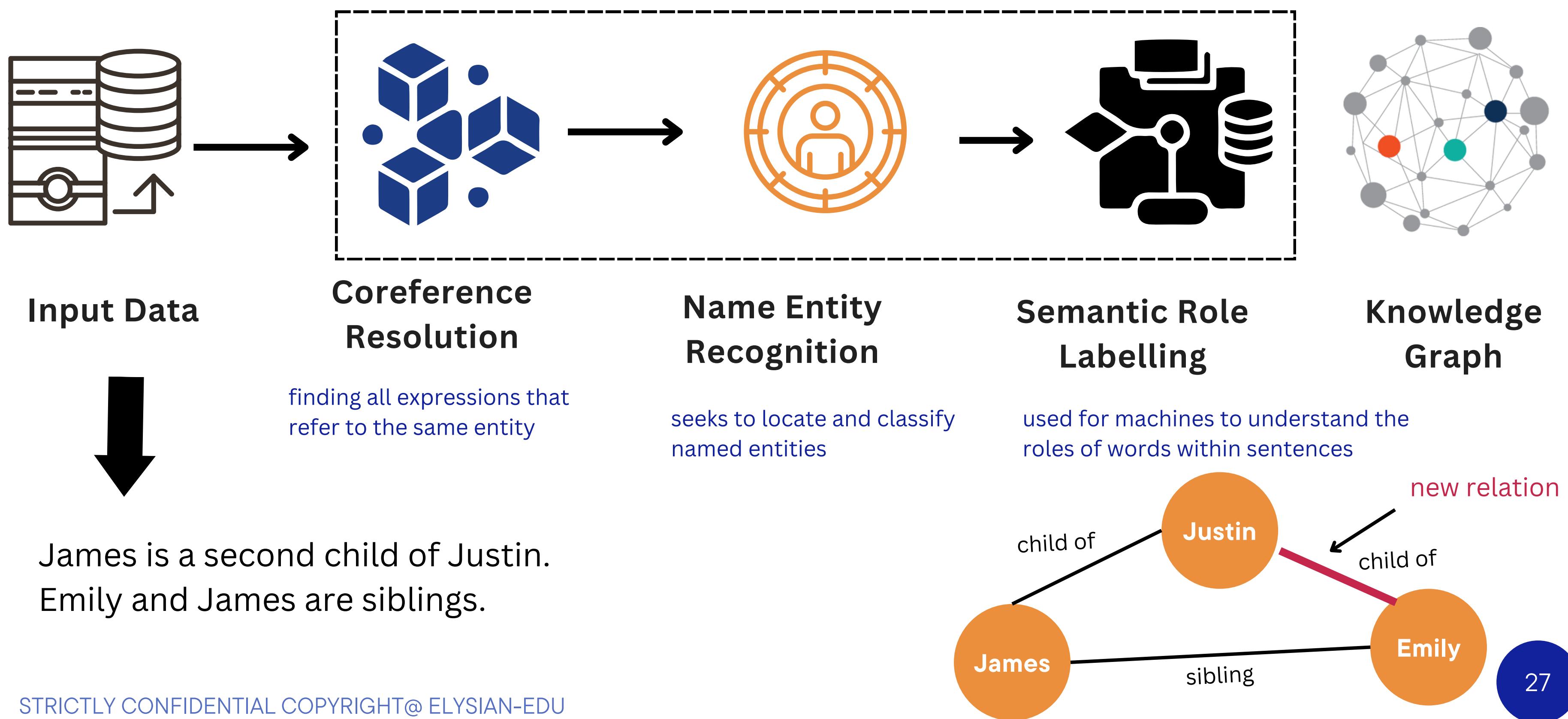
Word Embedding



Sentence Embedding



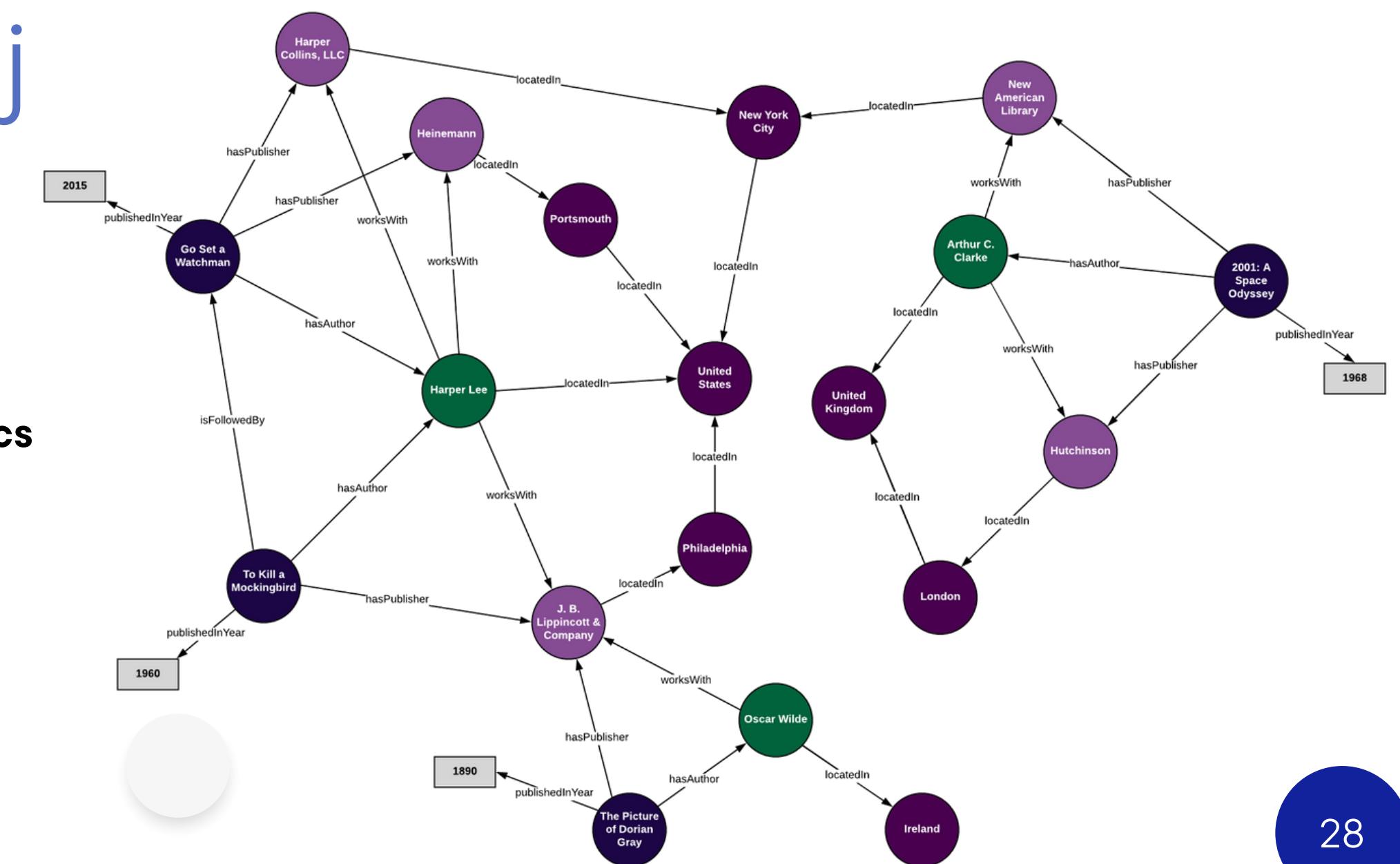
Data Extraction Pipelines using Natural Language Processing



Case Study : Knowledge graph using Neo4j

Here we go.....

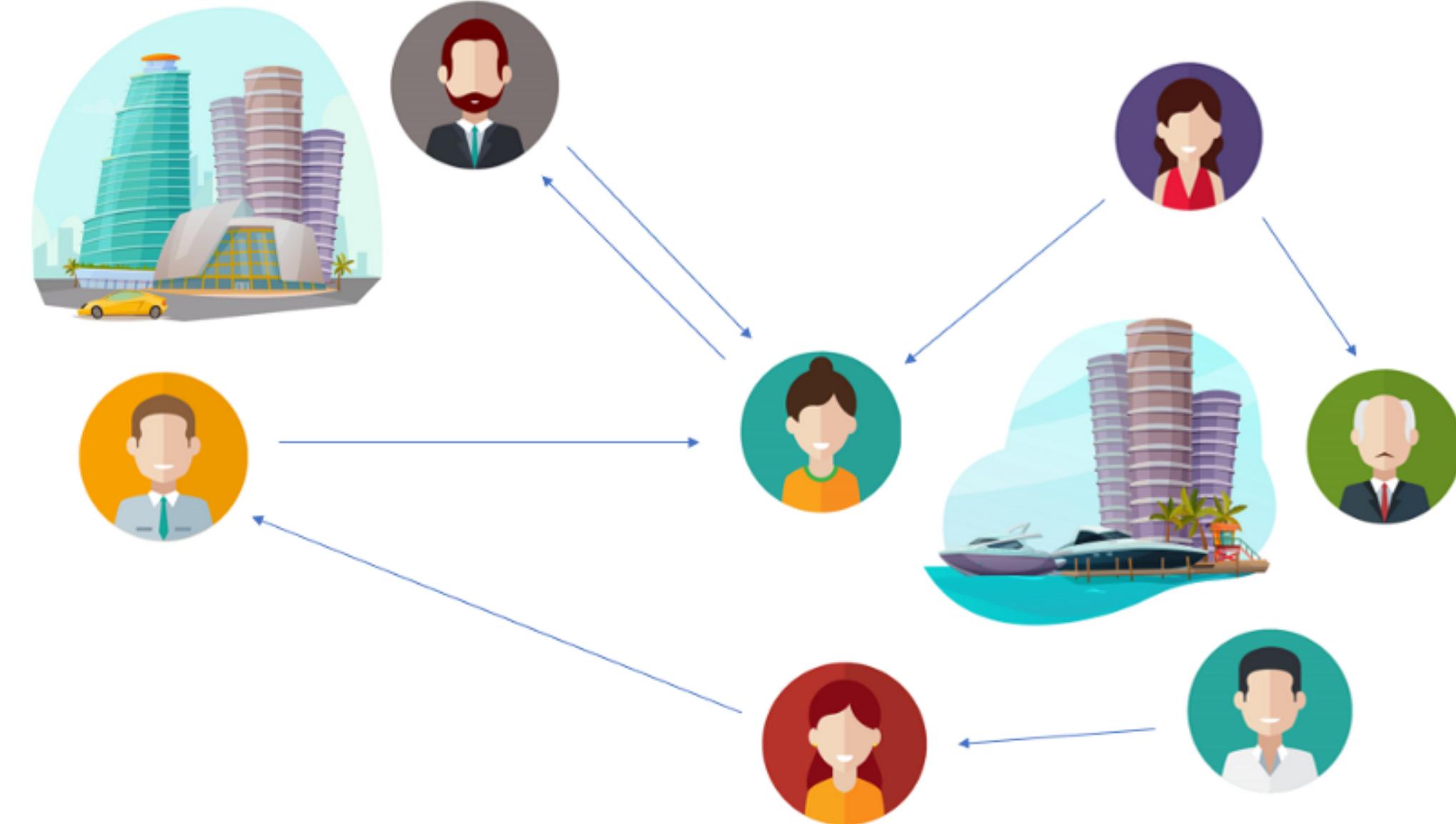
A Machine Learning Innovation in Predictive Analytics



Calls Logs Dataset

1. Data Domain

- There are some cities.
- There are some people, living in those cities.
- People call each other.

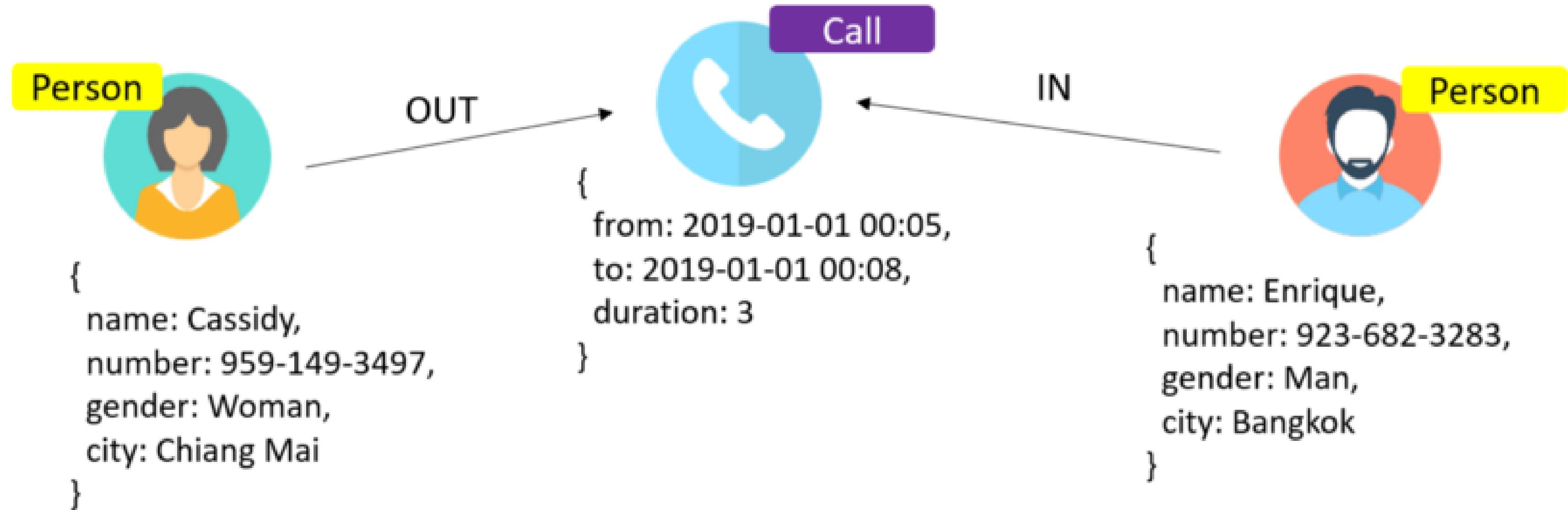


2. Data Modelling

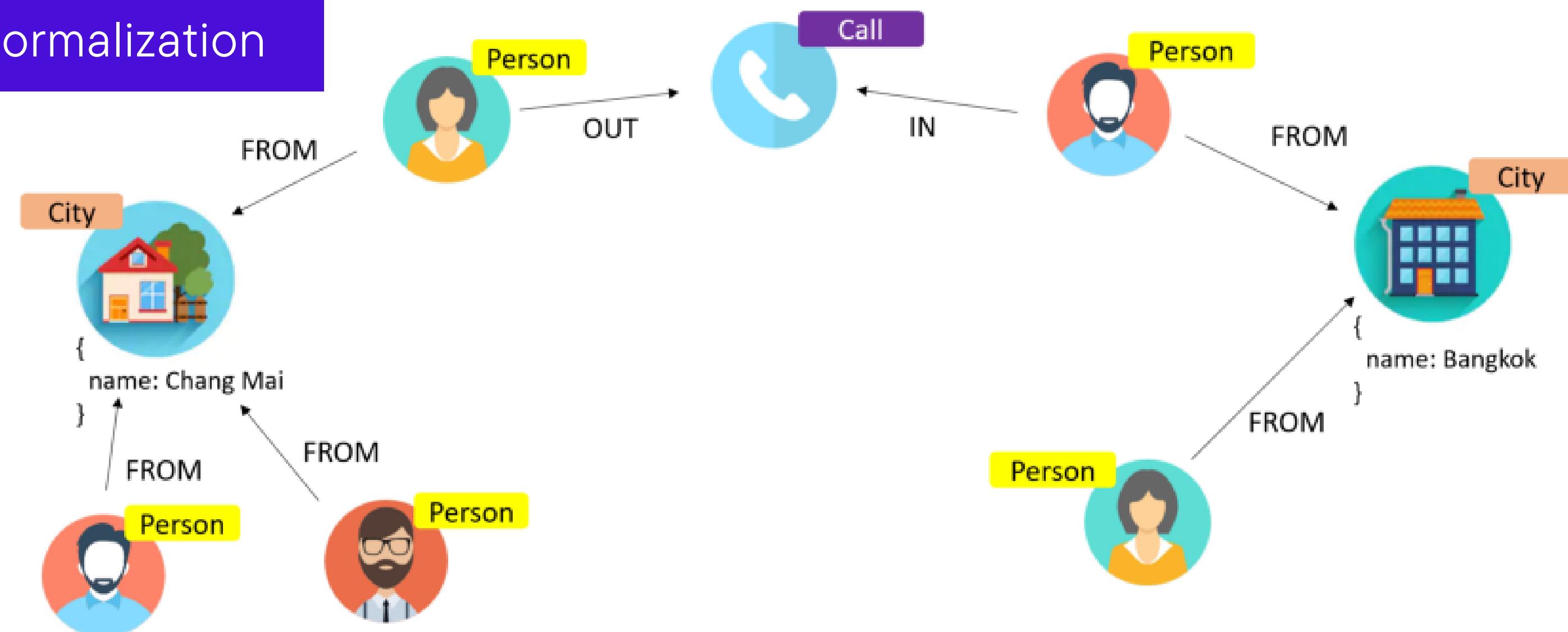
| | | | | | | | | | |
|------------------|------------------|------------|-------|---------|--------------|---------|-----|---------|--------------|
| 2019-01-01T06:04 | 2019-01-01T06:06 | Chiang Mai | Woman | Cassidy | 959-149-3497 | Bangkok | Man | Enrique | 923-682-3283 |
|------------------|------------------|------------|-------|---------|--------------|---------|-----|---------|--------------|

Cassidy made a call to Enrique.

3. Mapping the Relationship



4. Normalization



5. Finding insights

1. How many calls were missed in May?
 2. Find a city with the lowest number of internal city calls?
 3. On date 25 of April, find the woman who has the least total duration of conversations?
 4. How many pairs of people, where persons called to each other?

Graph Representation Tools



NetworkX

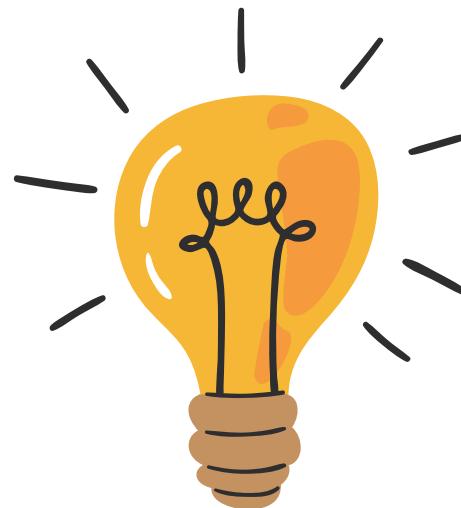


Conclusion



Today topic covers the following tips:

- Data vs Information vs Insights
- Data Science LifeCycle
- Challenges in DataScience/AI projects
- Predictive Models
 - Supervised Methods (Regression, Classification)
 - Unsupervised Methods (Clustering, Anomaly Detection)
- Graph Representation
 - CNN vs GNN
 - Data Mining with Knowledge Graph
- Use cases



Checkout the references for more details

THANK YOU!



FACEBOOK

Thet Su Win / Elysian Edu



LINKEDIN

Thet Su Win

