



Universidad
Politécnica
de Cartagena

MIEMBRO DE



EUROPEAN
UNIVERSITY OF
TECHNOLOGY



DATA MANAGEMENT, ANALYTICAL MODELS AND AI

Fernando Illescas Martínez



Table of Contents

1. Introduction to Big Data

1.1 Definition of Big Data and its importance

2. Fundamentals of Data Management

2.1 Types of Data

2.2 Data Storage

2.3 Data Processing

3. Introduction to Artificial Intelligence

3.1 Definition and Historical Evolution

3.2 Applications in Everyday Life

4. Analytical Models and Their Relationship with AI +

4.1 Analytical Models

4.2 Traditional Statistics vs Machine Learning

4.3 Machine Learning Algorithms

5. Supervised Learning vs Unsupervised Learning +

5.1 Differences and Applications

5.2 Examples of Algorithms in Each Category

6. Model Evaluation

6.1 Evaluation Metrics and Cross-Validation

7. Deep Learning and Neural Networks

7.1 Basic Concepts of Neural Networks

7.2 Examples of Applications in Computer Vision and Natural Language Processing

8. Case Studies

9. Ethical and Privacy Challenges

10. Learning Resources

1. Introduction to Big Data

1.1 Definition of Big Data and its importance

Big Data refers to extremely large and complex datasets that require special tools and techniques for processing and analysis. These data can be structured, semi-structured, or unstructured and come from various sources, such as social networks, sensors, online transactions, among others.



1.1 Definition of Big Data and its importance

Some examples of applications include:

- Sales data analysis to enhance a company's marketing strategy.
- Air quality monitoring through sensors to improve public health.
- Traffic data analysis to enhance urban planning.



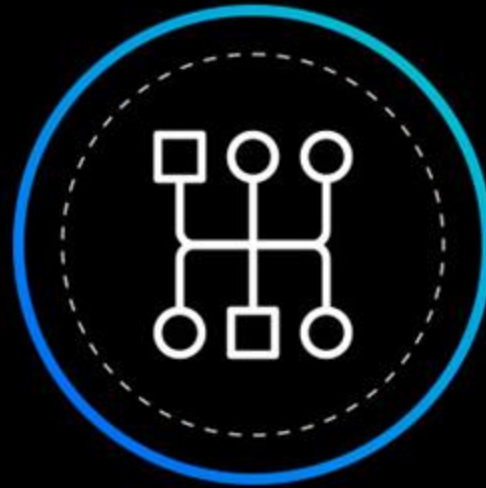
2. Fundamentals of Data Management

2.1 Types of Data



Structured
Data

vs



Semistructured
Data

vs



Unstructured
Data

2.2 Data Storage

TYPES OF STORAGE

- Cloud Storage
- Physical Storage Devices

STORAGE CONSIDERATIONS

- Storage Capacity
- Access Speed
- Security



2.3 Data Processing

Data processing is a fundamental stage in the analysis of large volumes of information. It involves the transformation of raw data into useful information for decision-making. This process comprises several key components:

- Data Cleaning
- Data Integration
- Data Transformation
- Data Loading



2.3 Data Processing

Types of Data Processing

There are different types of data processing, depending on the needs and objectives of the organization. Among the most common are:

- Real-time
- Batch
- In-Memory



3.

Introduction to Artificial Intelligence



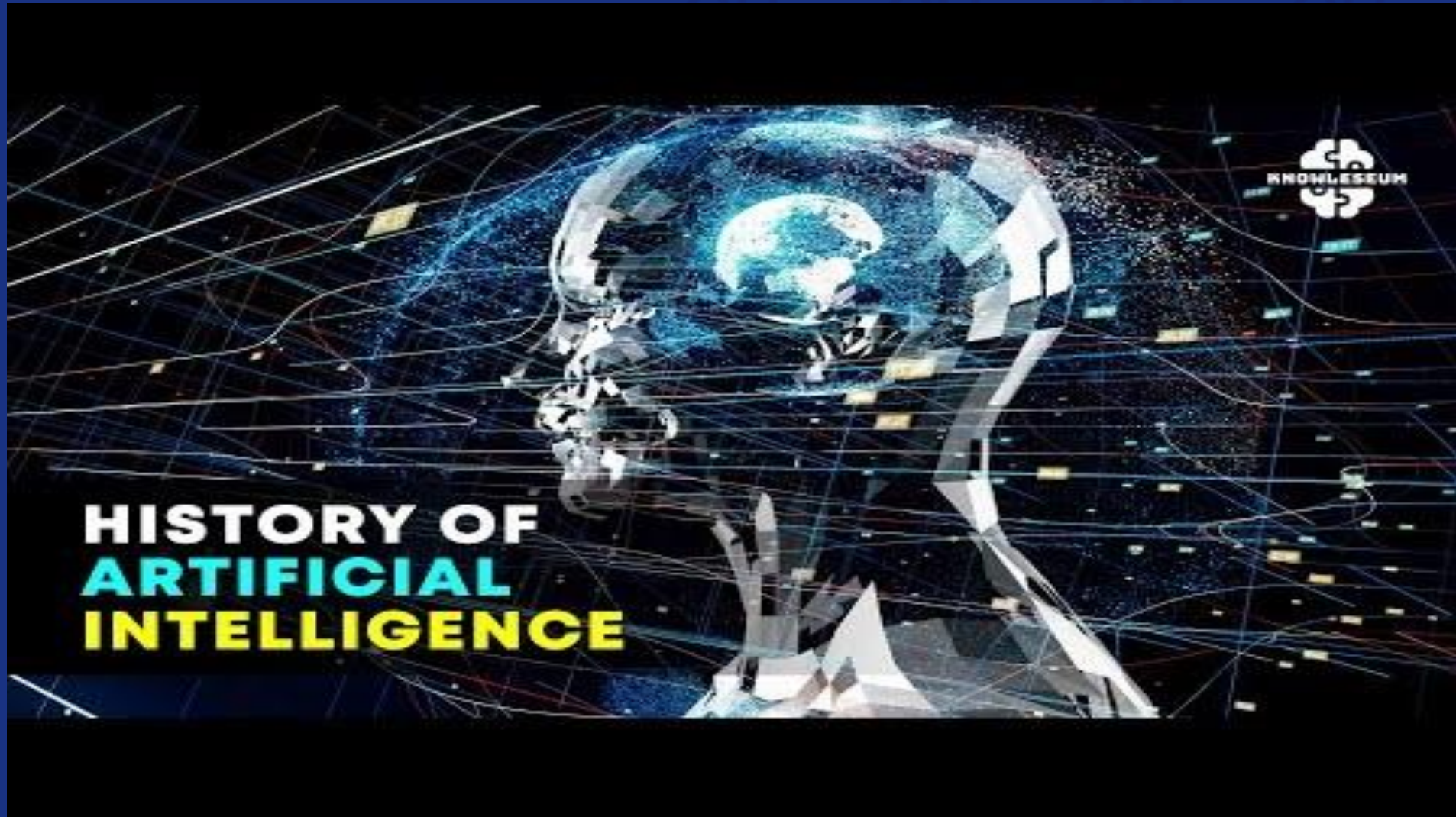
Universidad
Politécnica
de Cartagena

MIEMBRO DE



EUROPEAN
UNIVERSITY OF
TECHNOLOGY

3.1 Definition and Historical Evolution



3.2 Applications in Everyday Life

Artificial Intelligence (AI) is present in many aspects of our daily life, from virtual assistants on our phones to recommendation systems on streaming platforms. Some common applications include:

- Virtual assistants
- Social media platforms
- Facial recognition
- Recommendation systems on streaming platforms



4. Analytical Models and Their Relationship with AI

4.1 Analytical Models

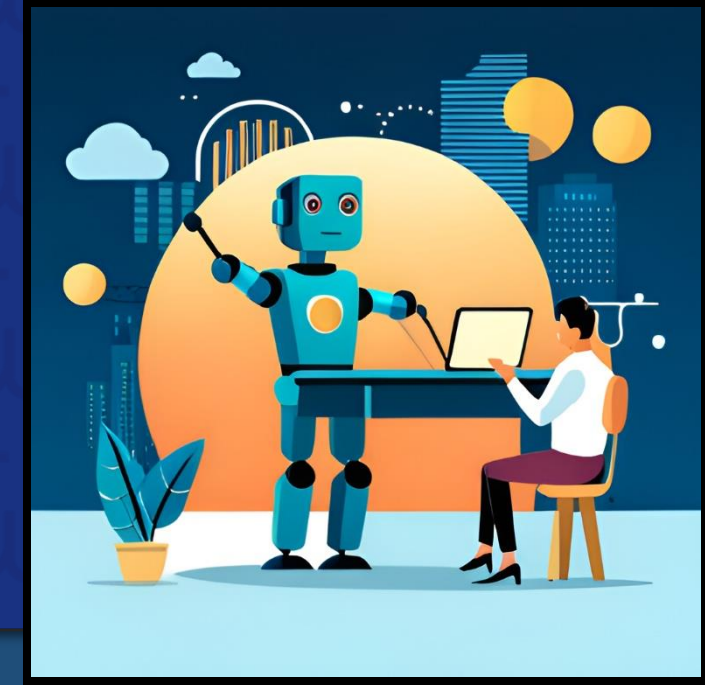
- **Analytical models** are tools that enable companies to make data-driven informed decisions.
- **Analytical models** are used in a variety of fields, including finance, marketing, and data science



4.2 Traditional Statistics vs Machine Learning



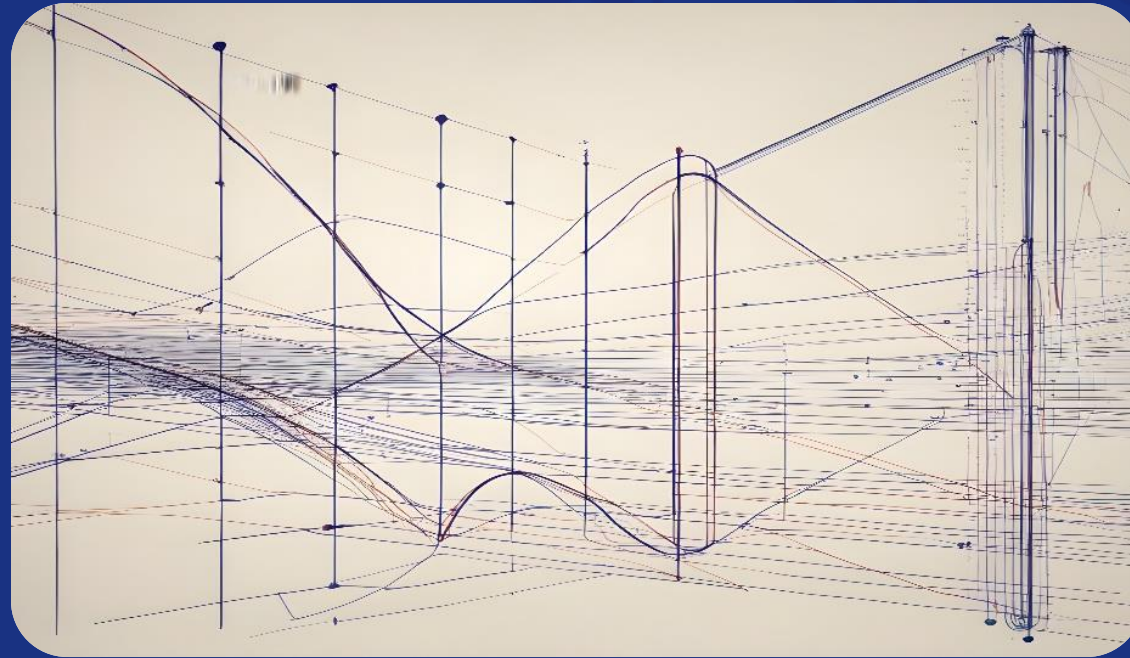
Traditional statistics are based on mathematical theory and are used to analyze data that conforms to a known distribution. They are used to describe and summarize data and to make inferences about a population from a sample.



Machine learning is a branch of artificial intelligence that focuses on building algorithms that can learn from data. It is used for making predictions and decisions based on patterns in data.

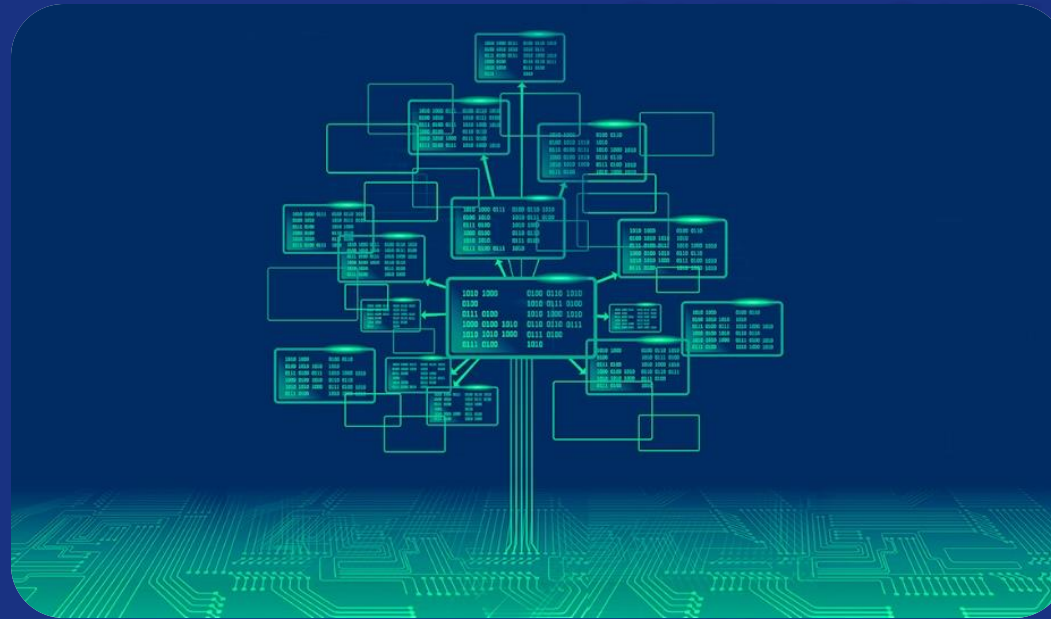
4.3 Machine Learning Algorithms

Linear regression is a supervised learning algorithm used to predict a continuous variable based on one or more independent variables. It is useful for modeling linear relationships between variables.



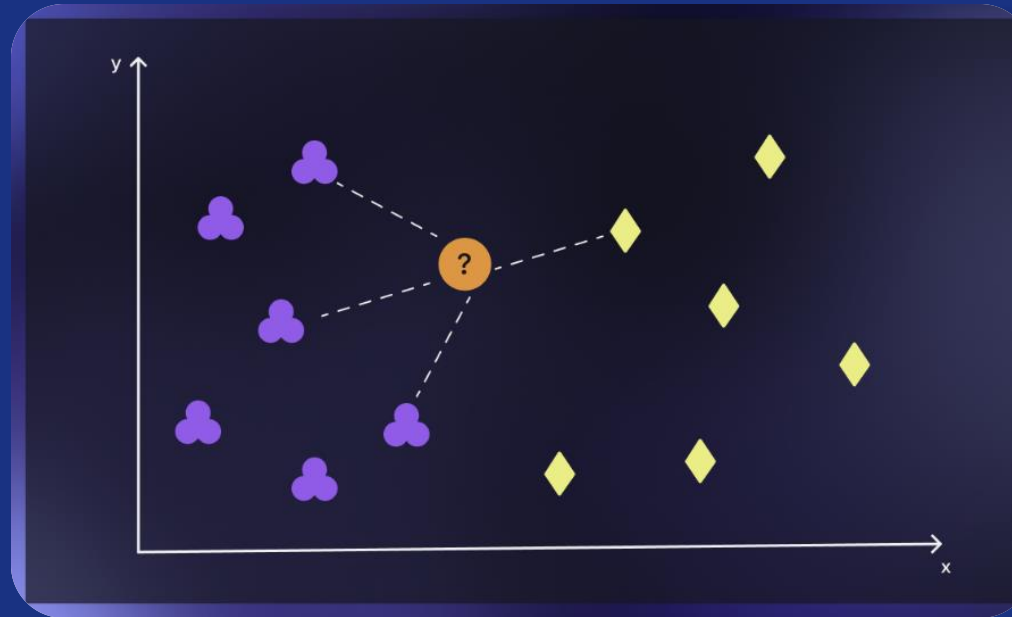
4.3 Machine Learning Algorithms

Decision trees are supervised learning algorithms used to classify data based on multiple features. They are useful for making decisions based on multiple criteria and for identifying patterns in data.



4.3 Machine Learning Algorithms

The **K-Nearest Neighbors** algorithm is a supervised learning algorithm used to classify data based on the characteristics of the nearest neighbors. It is useful for identifying patterns and for making predictions based on historical data.



5. Supervised Learning vs Unsupervised Learning

5.1 Differences and Applications

Supervised learning refers to a type of algorithm that uses labeled data to predict outcomes. On the other hand, unsupervised learning refers to a type of algorithm that uses unlabeled data to discover patterns and relationships in the data.

Differences

SUPERVISED LEARNING

- Input data is labelled
- Uses training dataset
- Used for prediction
- Classification and regression

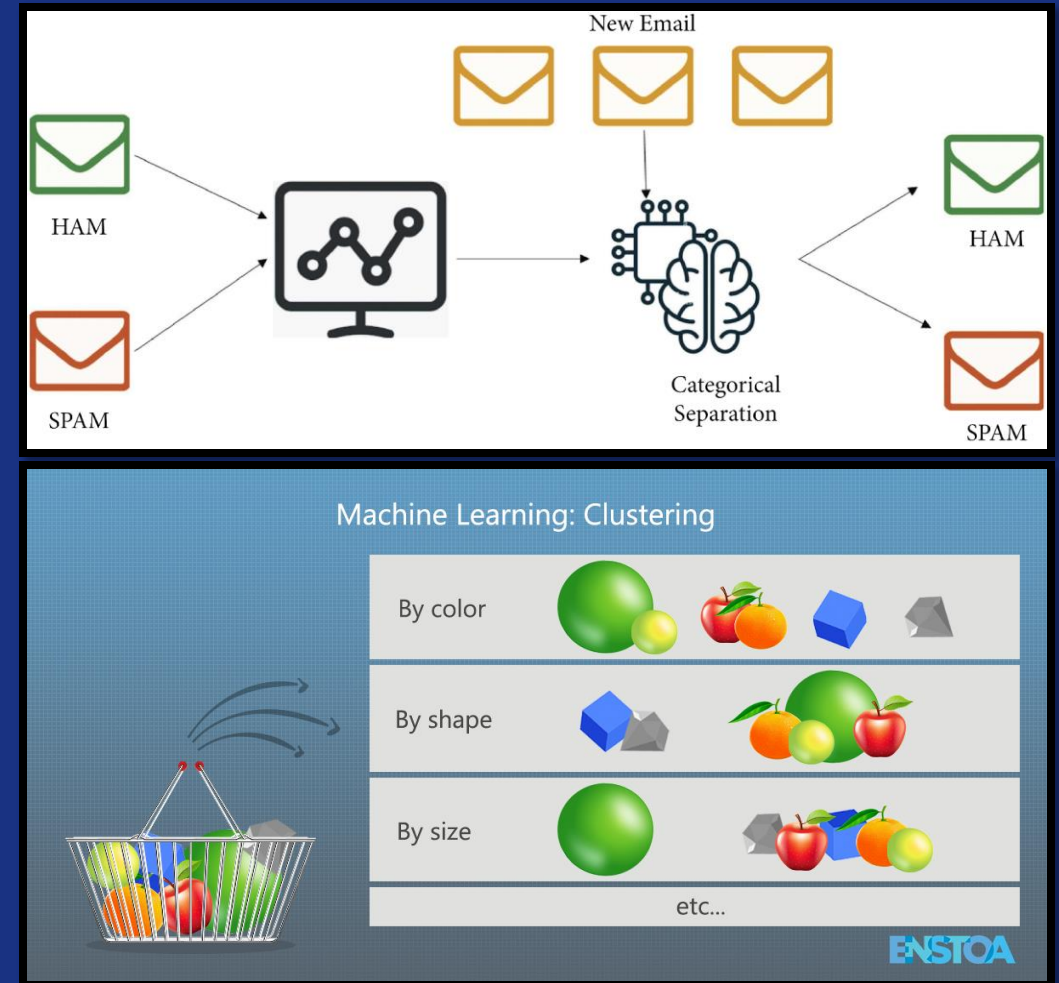
UNSUPERVISED LEARNING

- Input data is unlabelled
- Uses just input dataset
- Used for analysis
- Clustering, density estimation and dimensionality reduction

5.1 Differences and Applications

Applications

- Supervised learning is used in applications such as email classification, spam detection, and customer behavior prediction.
- Unsupervised learning is used in applications such as customer segmentation, anomaly detection, and data clustering.



5.2 Examples of Algorithms in Each Category

SUPERVISED LEARNING

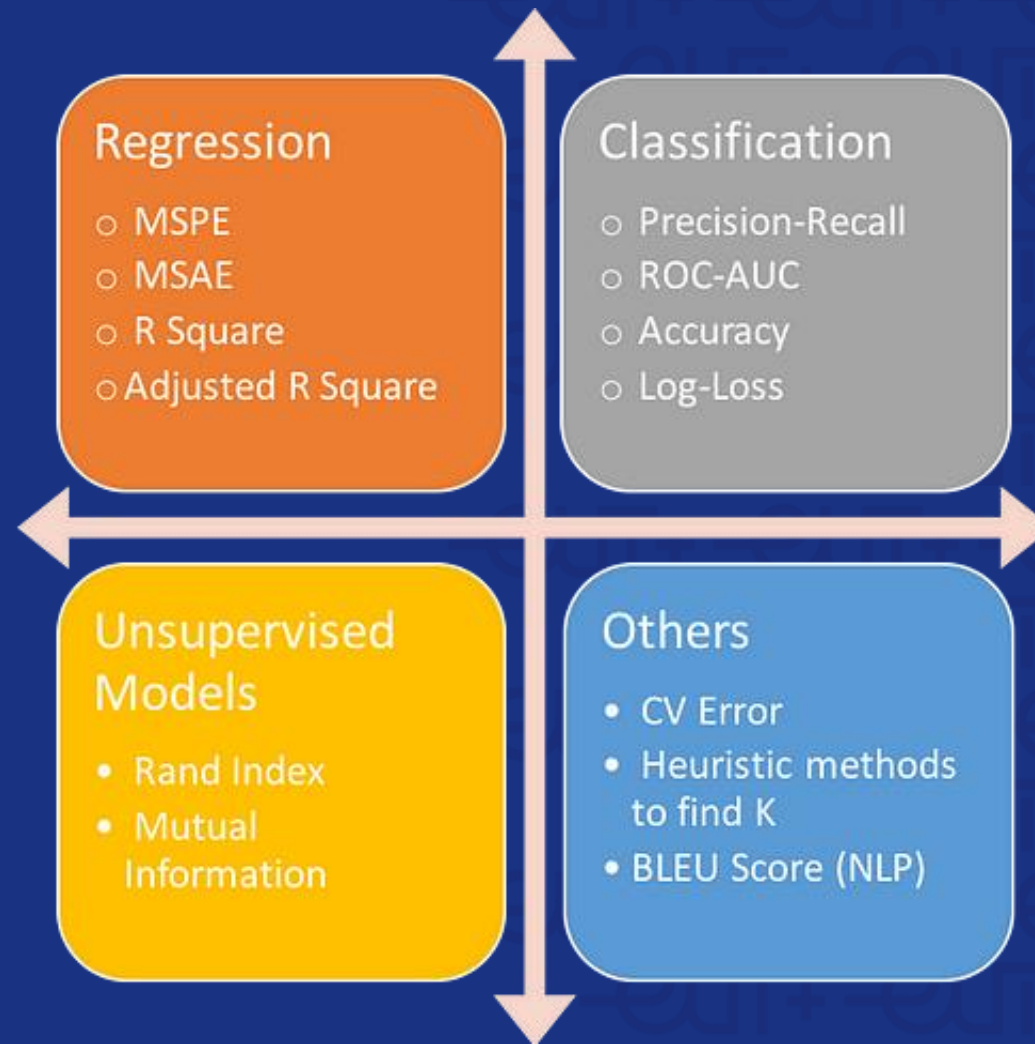
- Linear Regression
- Logistic Regression
- Support Vector Machines
- Random Forest
- Neural Networks

UNSUPERVISED LEARNING

- K-Means Clustering
- Principal Component Analysis (PCA)
- Hierarchical Clustering
- Association Rules
- t-Distributed Stochastic Neighbor Embedding (t-SNE)

6. Model Evaluation

6.1 Evaluation Metrics and Cross Validation



6.1 Evaluation Metrics and Cross Validation

Cross-validation is a technique used to assess the performance of a model on a dataset. It involves dividing the dataset into k subsets, where k is an integer. The model is then trained k times, using each of the subsets as a test set in one of the iterations and the remaining subsets as the training set. At the end of the k iterations, the results are averaged to obtain a more accurate measure of the model's performance.

4-fold validation ($k=4$)

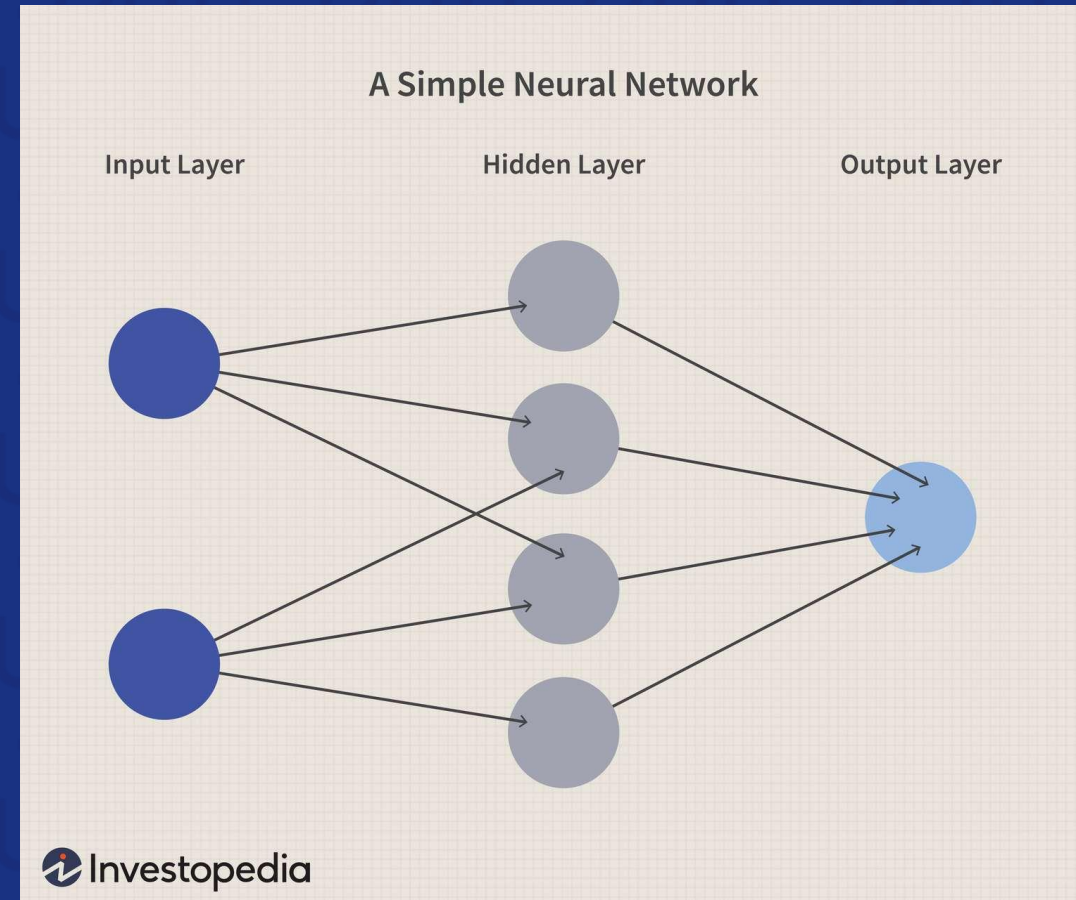


7. Deep Learning and Neural Networks

7.1 Basic Concepts of Neural Networks

What are Neural Networks?

Neural networks are a set of algorithms and mathematical models used to analyze and process large amounts of data. They are inspired by the functioning of the human brain and are used in various areas, such as computer vision, natural language processing, and decision-making.

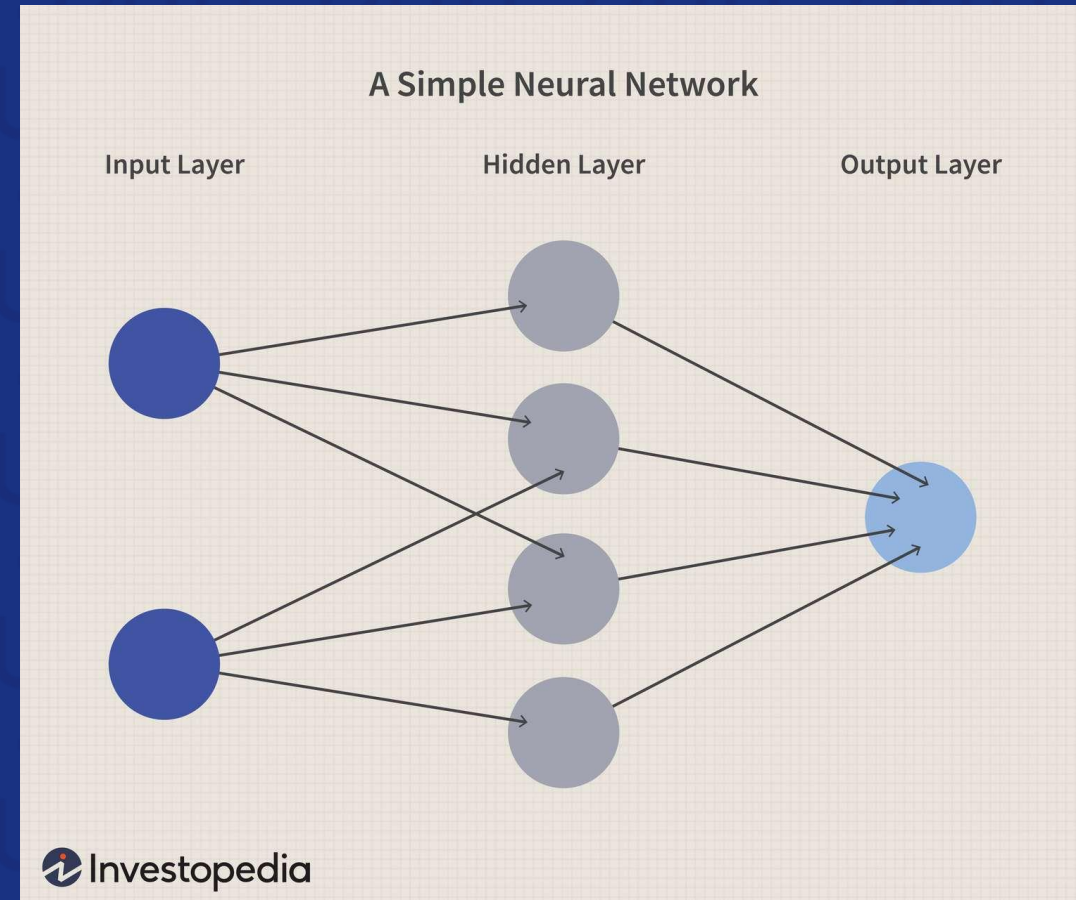


7.1 Basic Concepts of Neural Networks

Component of a Neural Network

A neural network consists of three main components: neurons, connections, and layers.

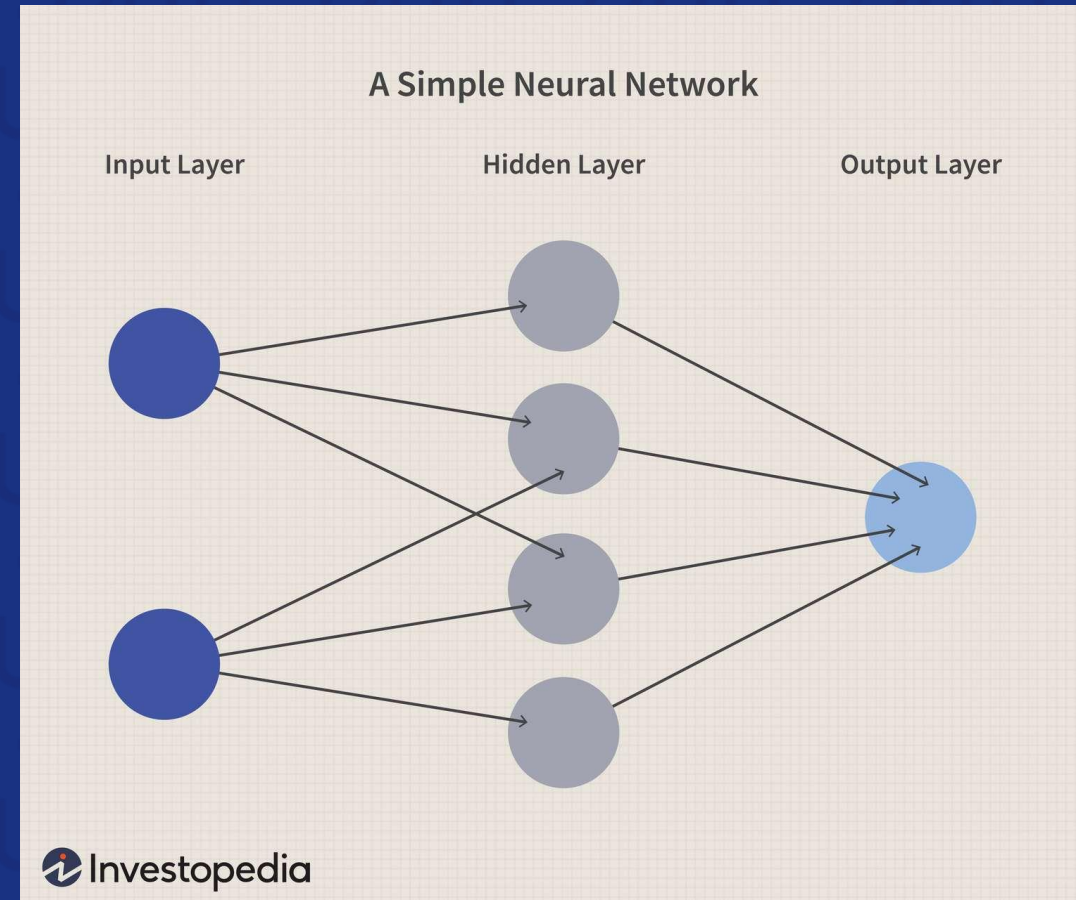
- Neurons are basic processing units that receive inputs and produce outputs.
- Connections are the links between neurons that transmit information.
- Layers are groups of neurons that process information in parallel.



7.1 Basic Concepts of Neural Networks

How a Neural Network Works

The operation of a neural network is based on learning through feedback. During the training process, the neural network receives a set of input data and produces an output. The output is compared to the desired output, and the error is calculated. The network adjusts its parameters to minimize the error and improve its accuracy.



7.2 Examples of Applications in Computer Vision and Natural Language Processing

COMPUTER VISION

- Object Detection
- Facial Recognition
- Image Segmentation



7.2 Examples of Applications in Computer Vision and Natural Language Processing

NATURAL LANGUAGE PROCESSING

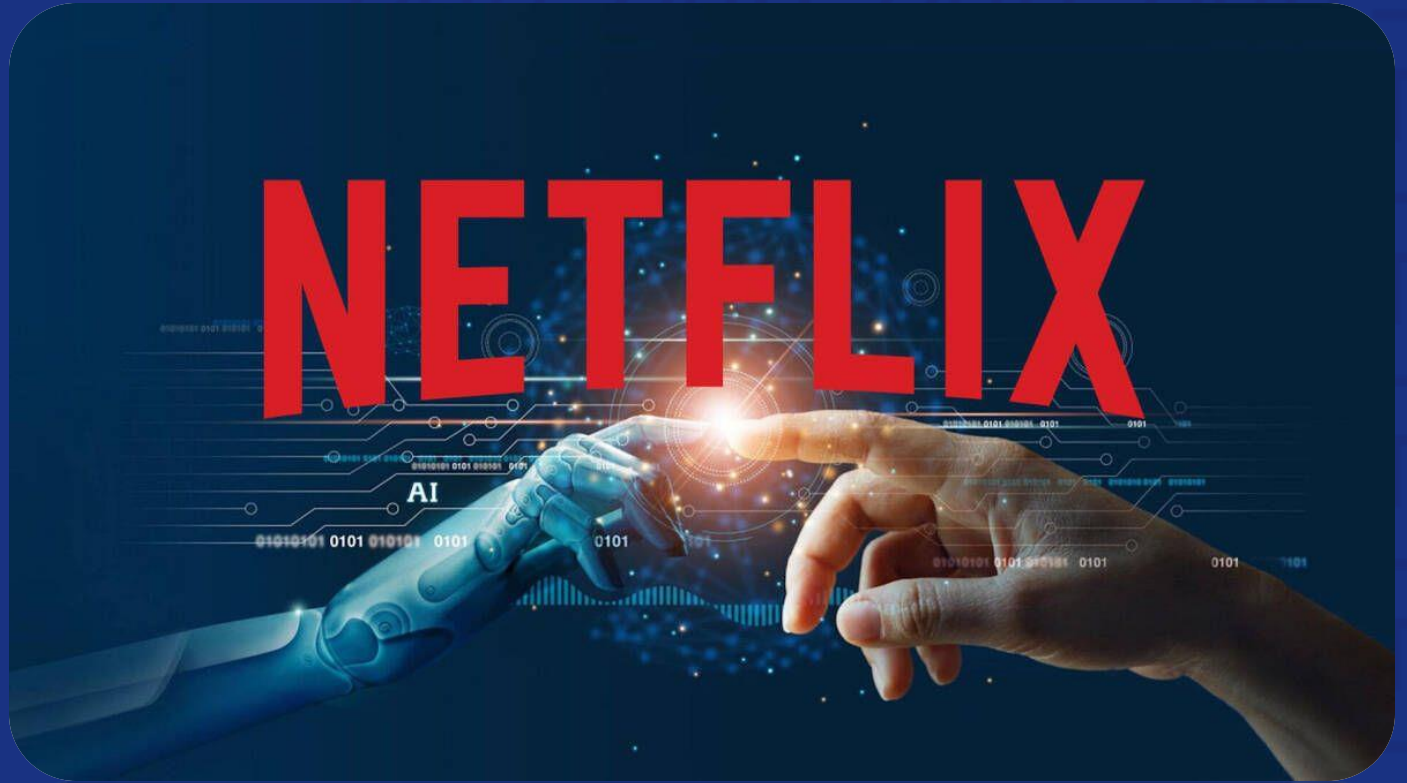
- Sentiment Analysis
- Speech Recognition
- Machine Translation



8. Case Studies

8 Case Studies

Netflix is one of the largest users of big data in the entertainment world. They use data analytics to personalize the viewing experience for users and provide highly relevant content recommendations.



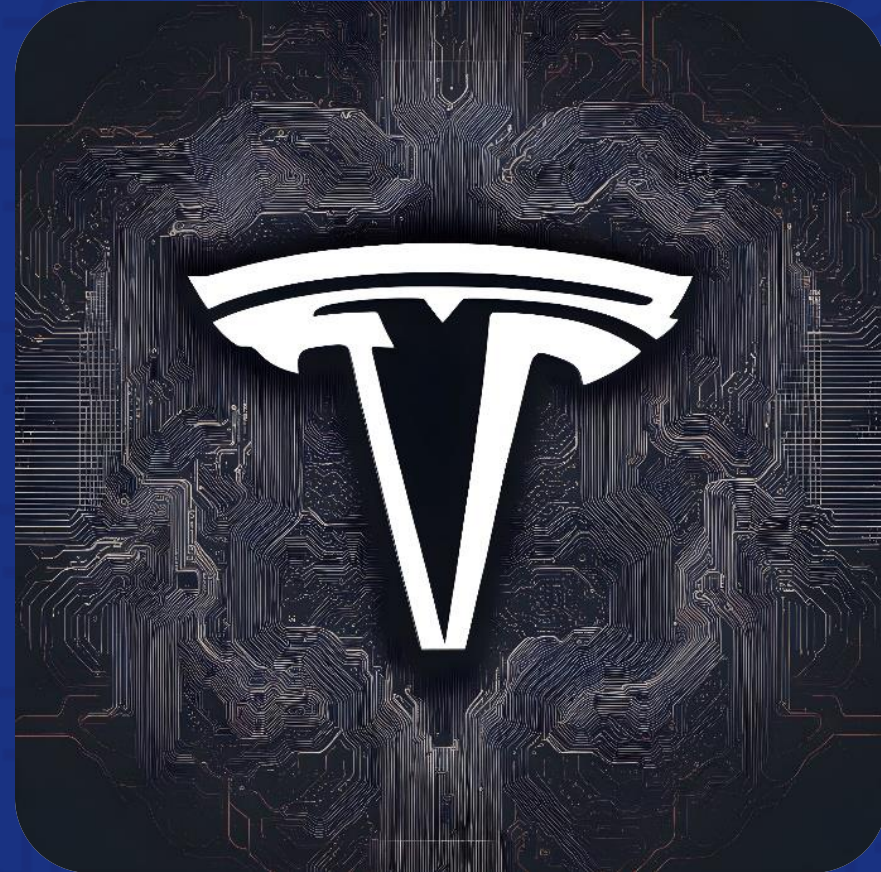
8 Case Studies

Amazon uses big data to enhance the user shopping experience and optimize the supply chain. They employ data analytics to predict market trends and adjust their prices accordingly.

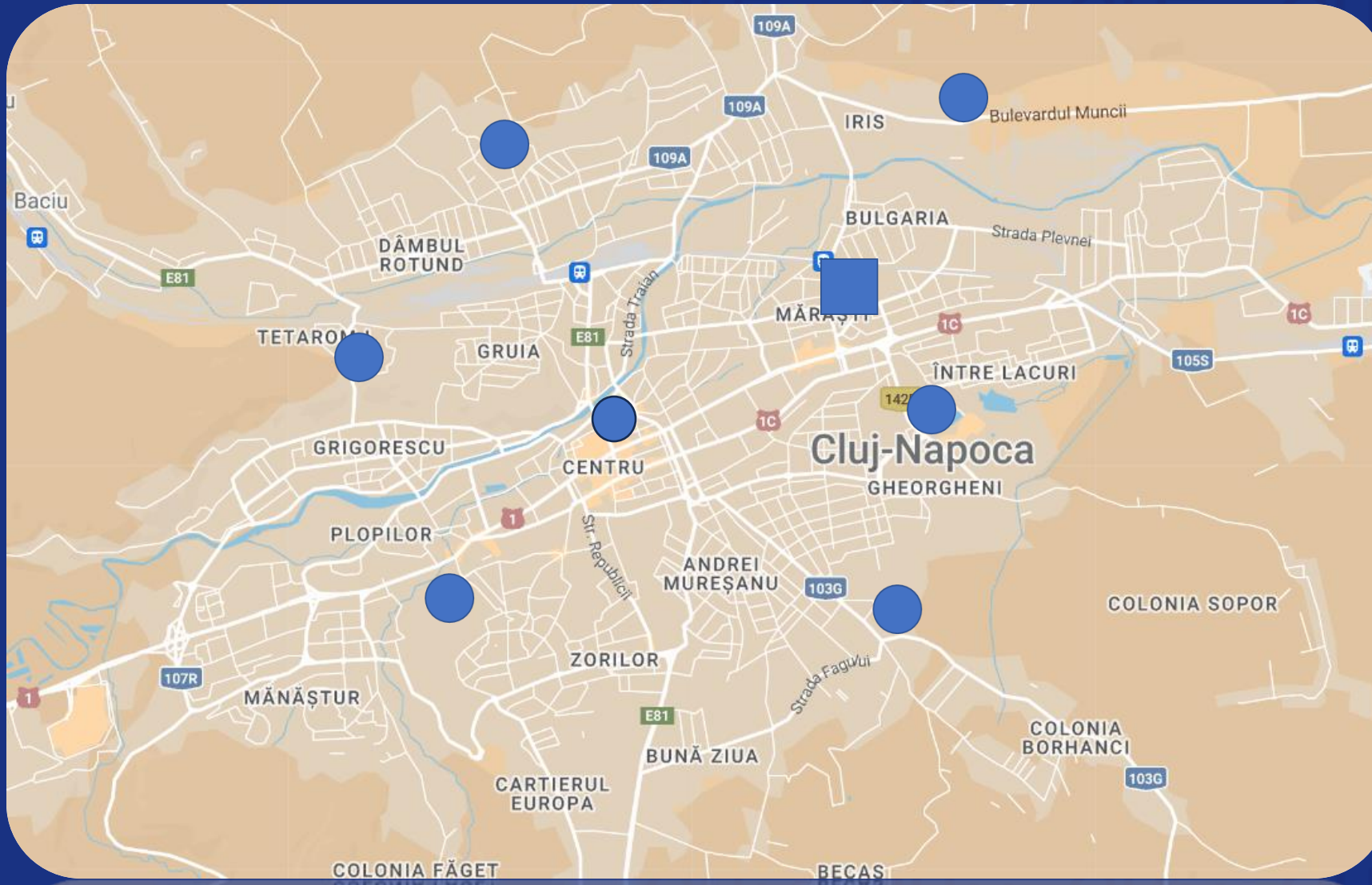


8 Case Studies

Tesla uses big data to enhance the efficiency of their electric vehicles and optimize their performance. They utilize data analytics to gather information about driver behavior .



8 Case Studies

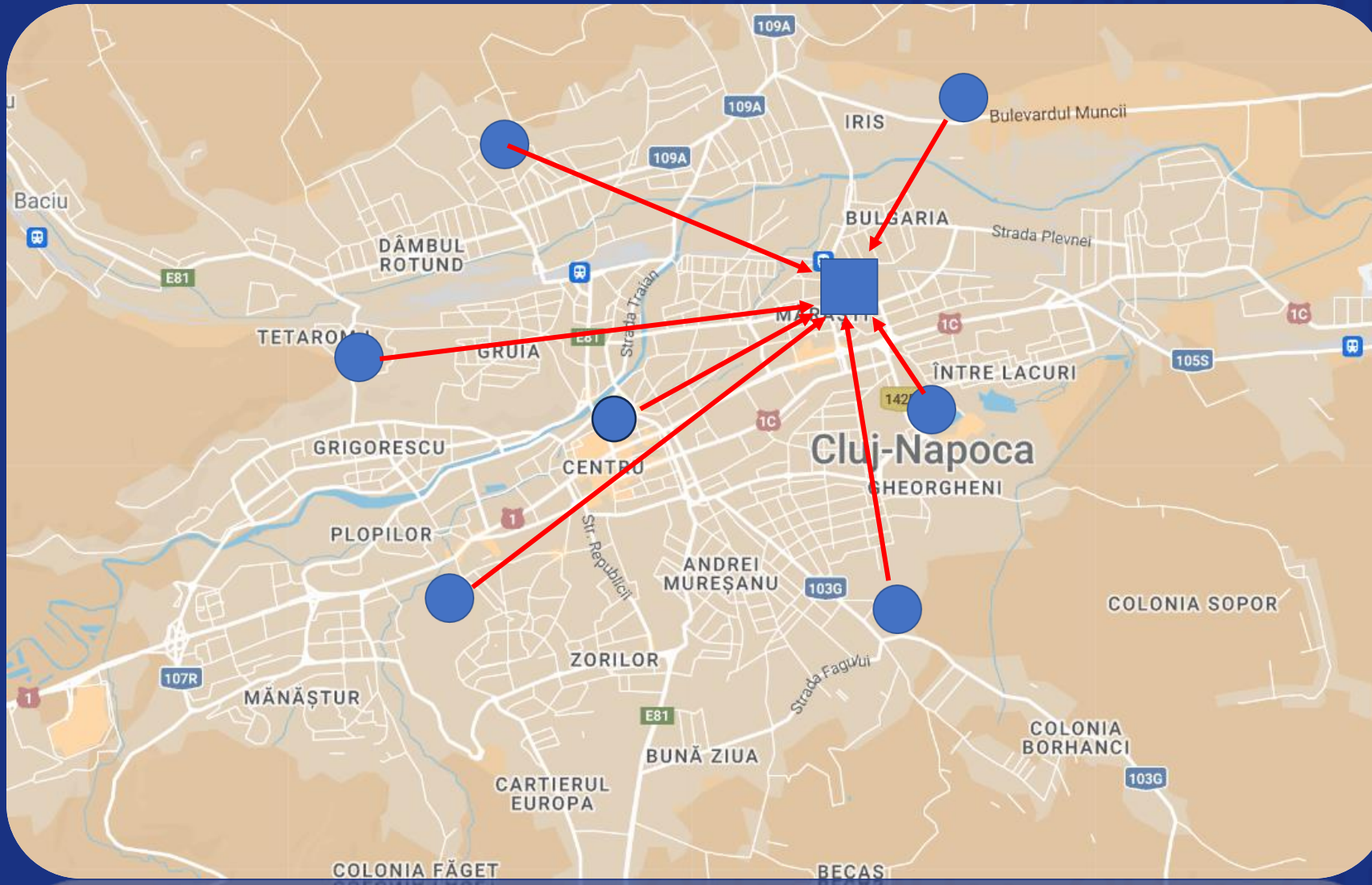


Node



GW

8 Case Studies

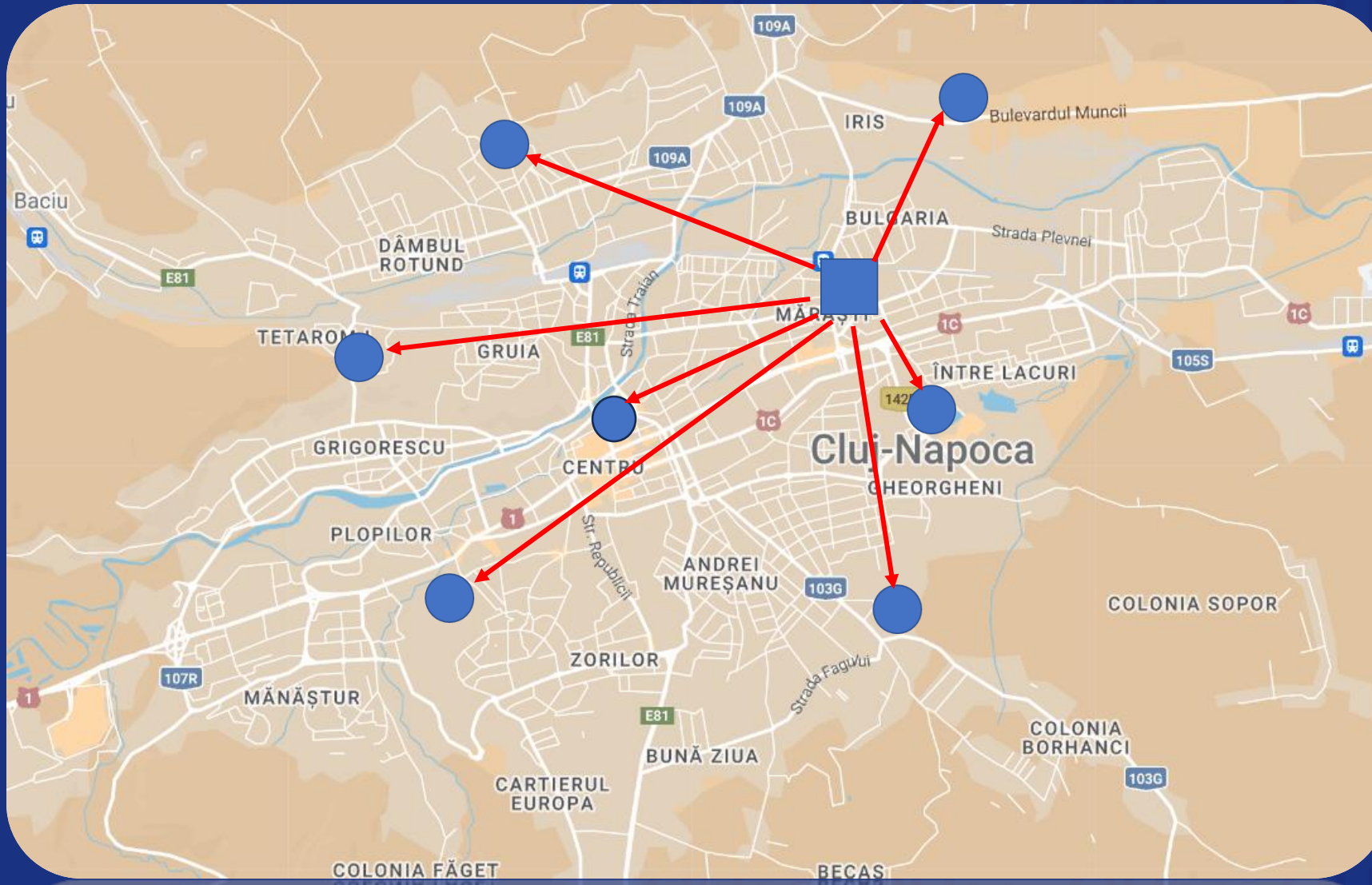


Node



GW

8 Case Studies

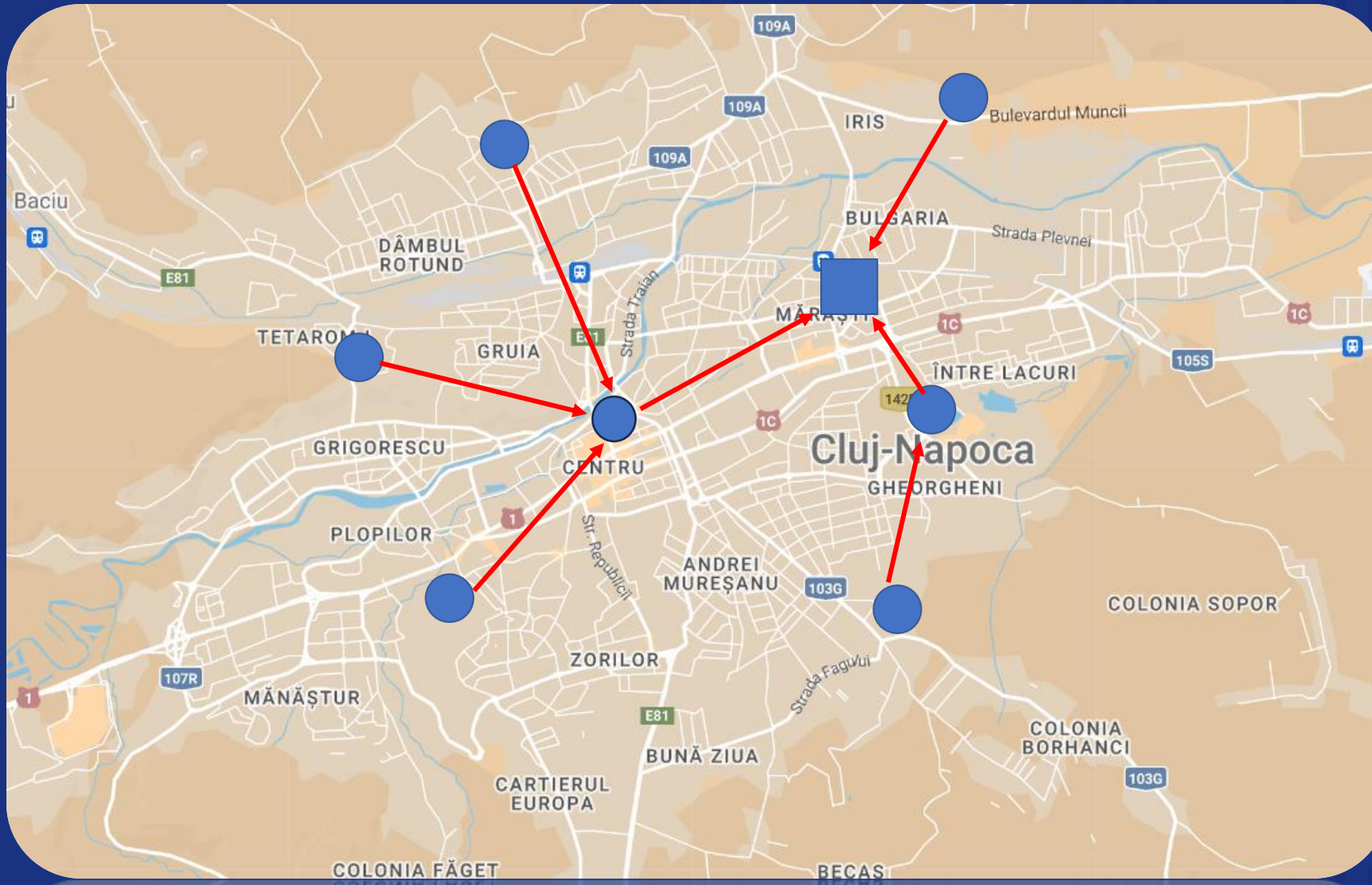


Node



GW

8 Case Studies

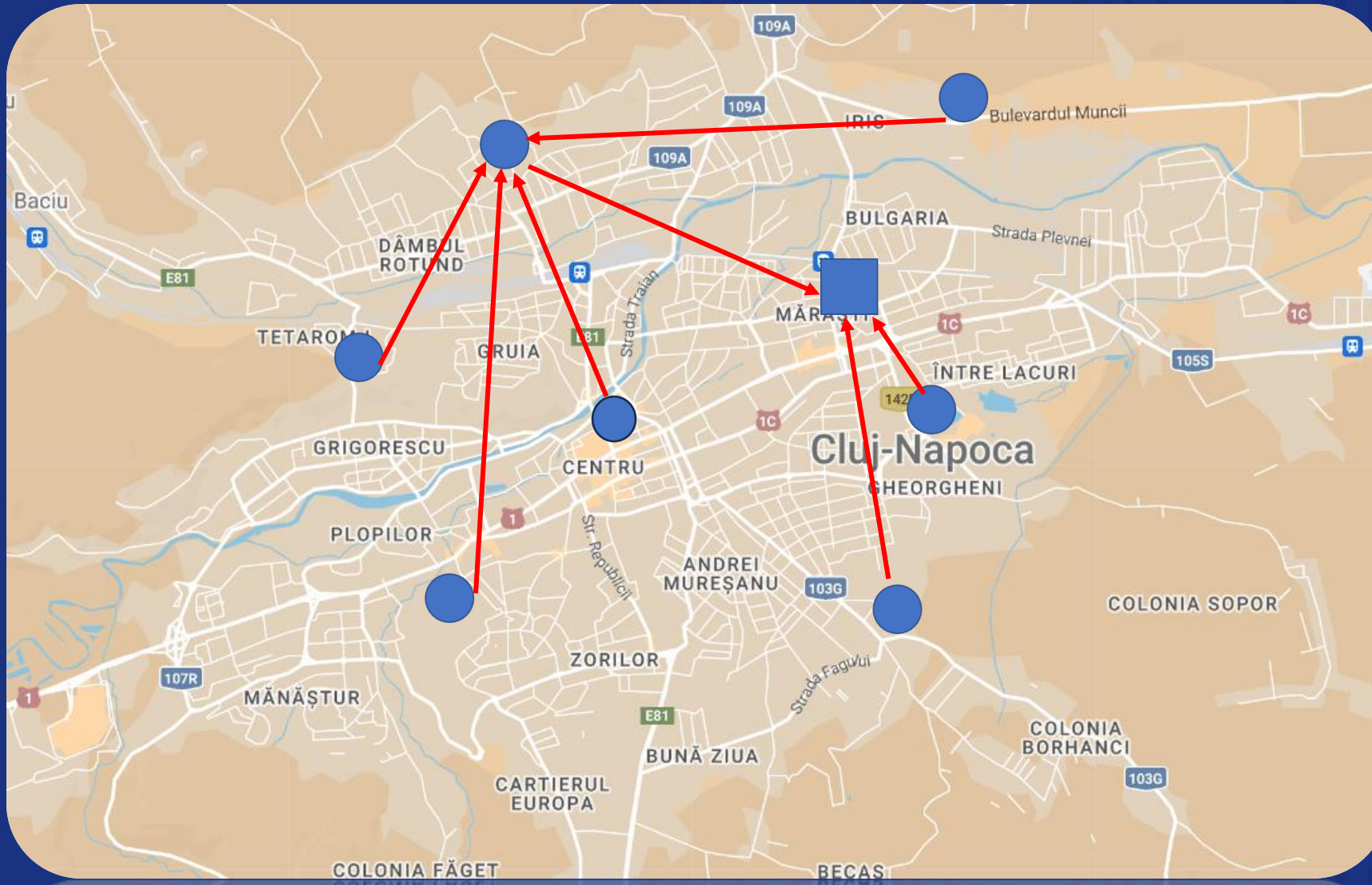


Node



GW

8 Case Studies

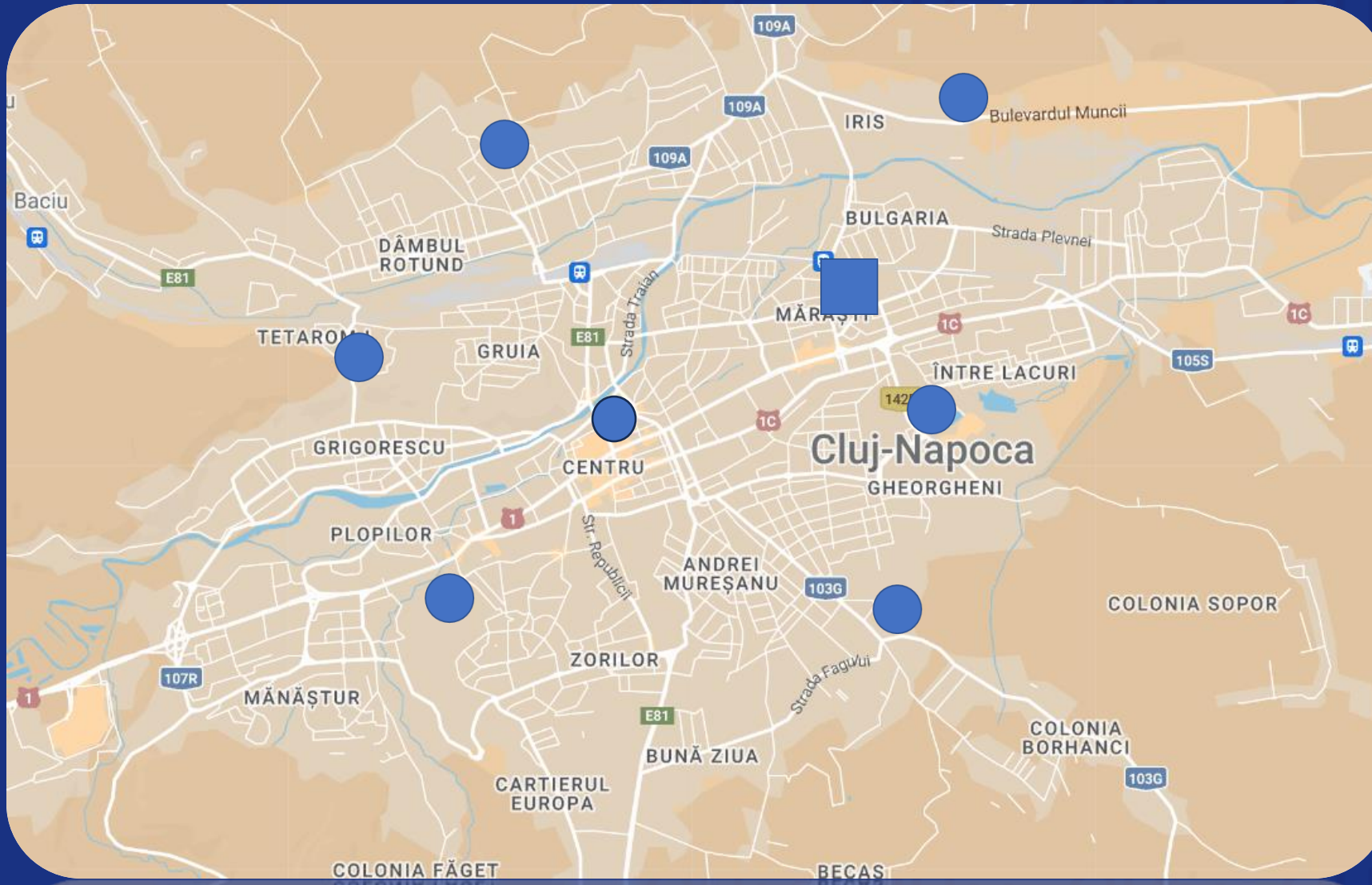


Node



GW

8 Case Studies

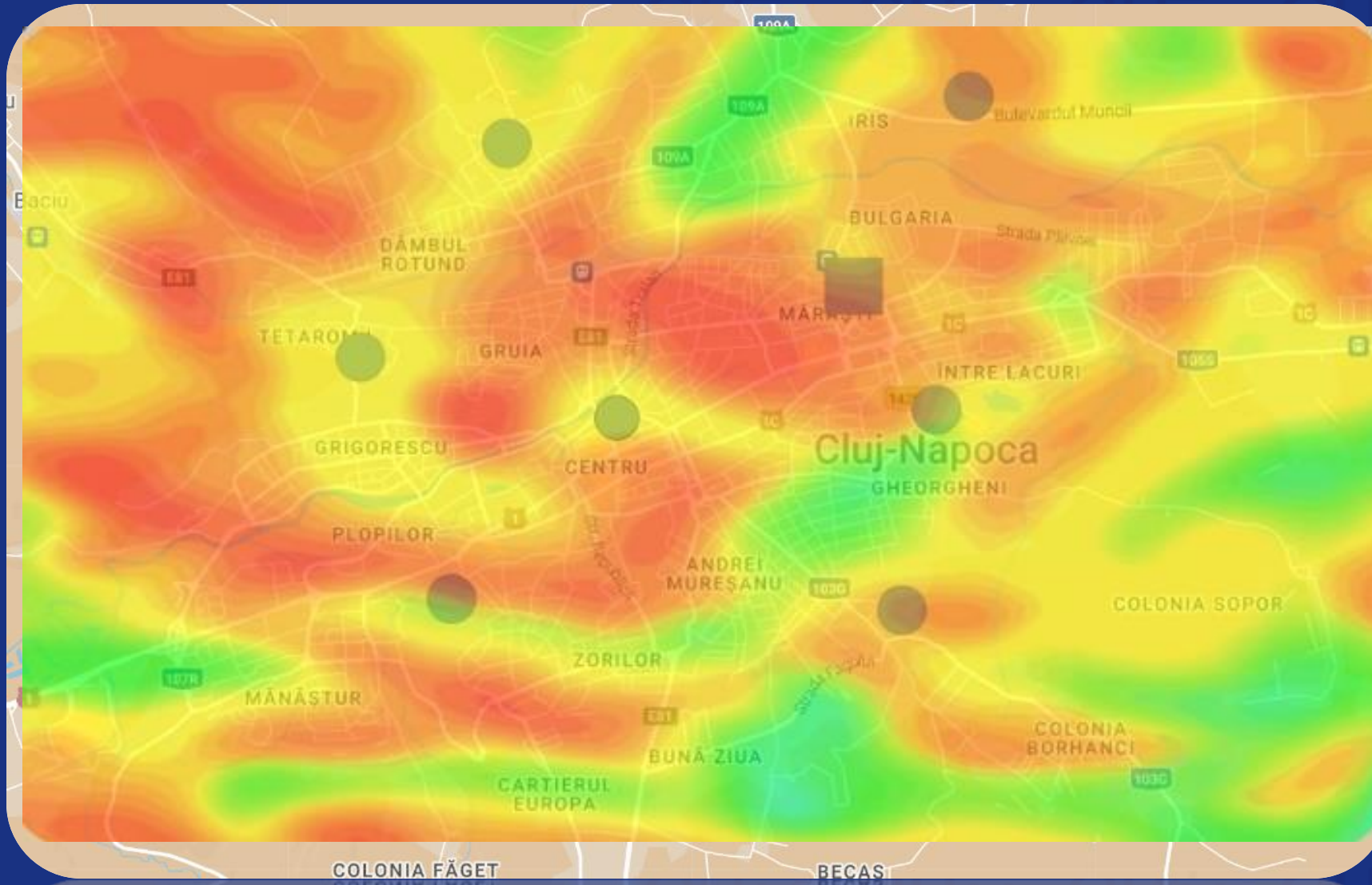


Node



GW

8 Case Studies



Node



GW

9. Ethical And Privacy Challenges

9 Ethical and Privacy Challenges

- **Data Privacy and Security:** Safeguarding sensitive data is paramount.
- **Algorithmic Fairness:** Address potential biases in algorithms.
- **Clear Policies:** Transparency in data management is crucial.
- **Informed Consent:** Users should be informed and provide consent.



10 Learning Resources

BOOKS

- "Data Science for Business", by Foster Provost and Tom Fawcett
- "Python Machine Learning", by Sebastian Raschka and Vahid Mirjalili
- "Deep Learning", by Ian Goodfellow, Yoshua Bengio and Aaron Courville

ONLINE COURSES

- Coursera - "Machine Learning", by Andrew Ng
- edX - "Data Science Essentials", by Microsoft
- Udemy - "Deep Learning A-Z", by Kirill Eremenko and Hadelin de Ponteves

COMMUNITIES AND FORUMS

- Kaggle - Community of data scientists and machine learning practitioners
- Stack Overflow - Programming and data analysis forum
- Reddit - Community for machine learning and deep learning



Universidad
Politécnica
de Cartagena

MIEMBRO DE



EUROPEAN
UNIVERSITY OF
TECHNOLOGY



THE END

DATA MANAGEMENT, ANALYTICAL MODELS AND AI

