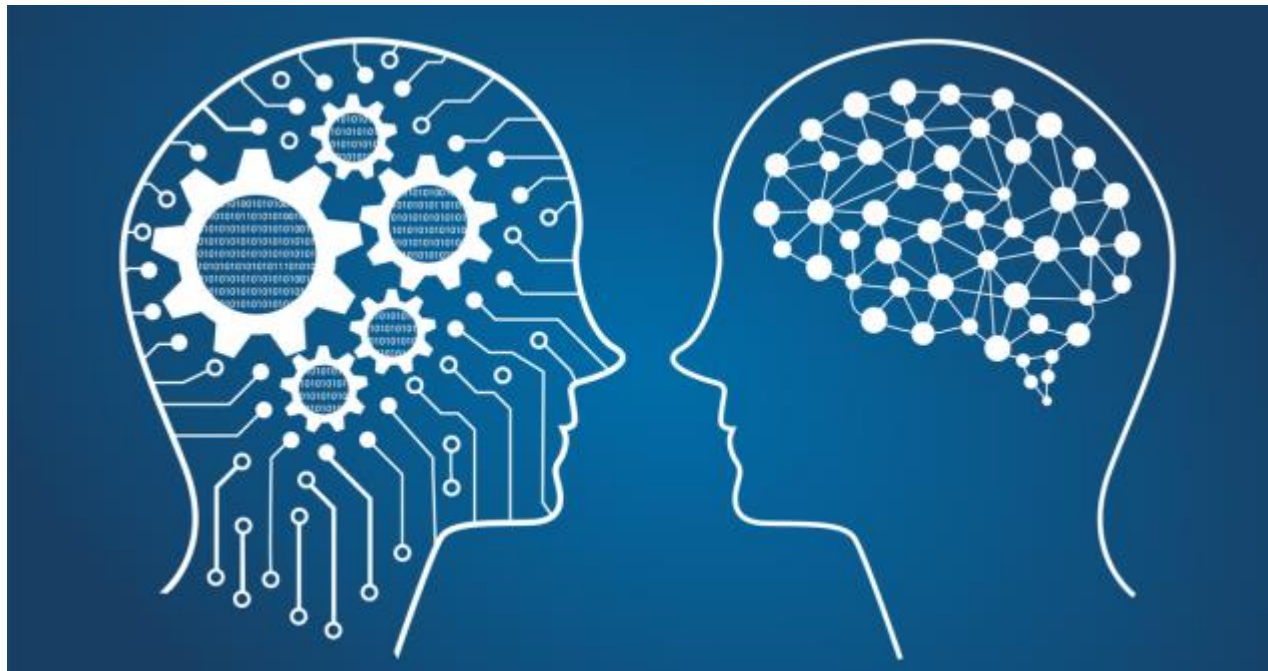


# Machine Learning

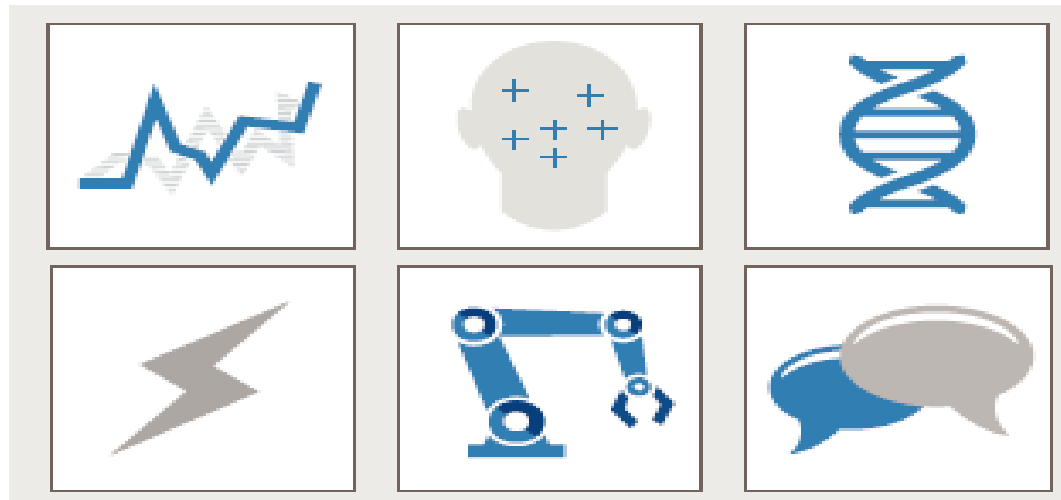
- What is Machine Learning?
- Real-World Application Areas
- Machine Learning Use Cases
- Machine Learning Challenges and Workflow
- Machine Learning techniques
- Applications in Infogain
- What next....

# What is Machine Learning?

- Teaches computers to do what comes naturally to humans and animals: learn from experience.
- Uses computational methods to “learn” information directly from data without relying on a predetermined equation as a model.
- Adaptively improves performance as the number of samples available for learning increases.



- With the rise in big data, machine learning has become particularly important for solving problems in areas like these:
  - Computational finance, for credit scoring and algorithmic trading
  - Image processing and computer vision, for face recognition, motion detection, and object detection
  - Computational biology, for tumor detection, drug discovery, and DNA sequencing
  - Energy production, for price and load forecasting
  - Automotive, aerospace, and manufacturing, for predictive maintenance
  - Natural language processing



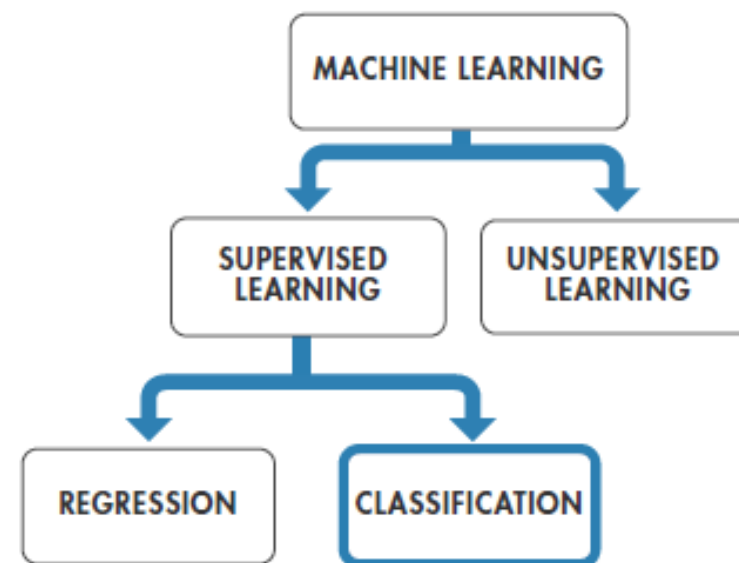
- Credit Risk and Fraud:
  - Citibank : collaboration with Portugal based fraud detection company Feedzai to identify and eliminate fraud in online and in-person banking by alerting the customer.
  - PayPal: uses machine learning to fight money laundering.
- Healthcare for Personalized Treatment
  - machine learning plays a vital role in finding what kind of genetic makers and genes respond to a particular treatment or medication.
- Focused Account Holder Targeting
  - machine learning is used to identify a group of home maker moms in Florida with huge social media presence to be their most influential and preferred banking customers in terms of referrals.
- Retail
  - Product Recommendations
  - Improved Customer Service

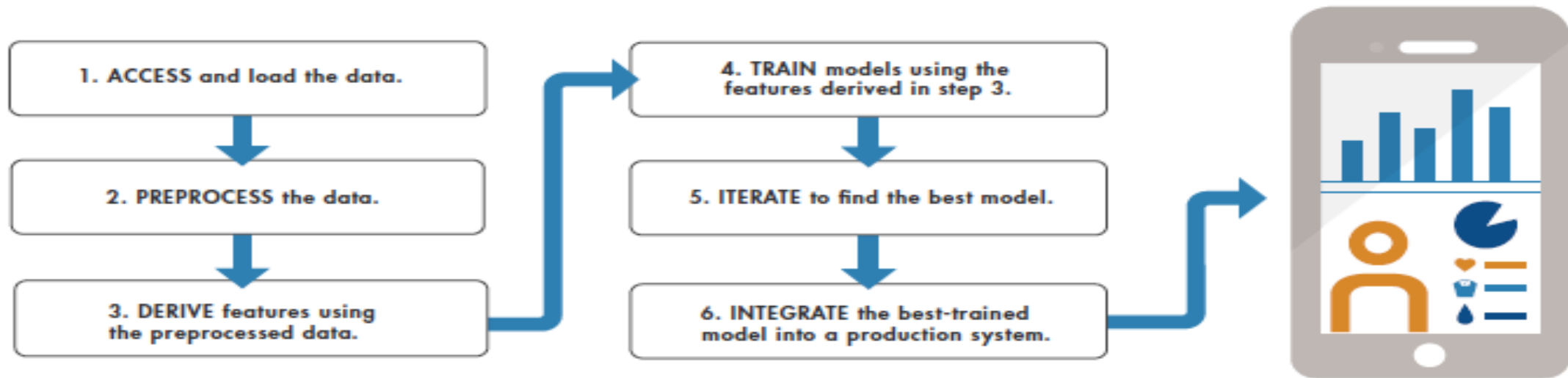
- Handling your data and finding the right model.
- Data comes in all shapes and sizes.
- Preprocessing data requires specialized knowledge and tools.
- Finding the best model to fit the data.



# Questions to Consider Before You Start

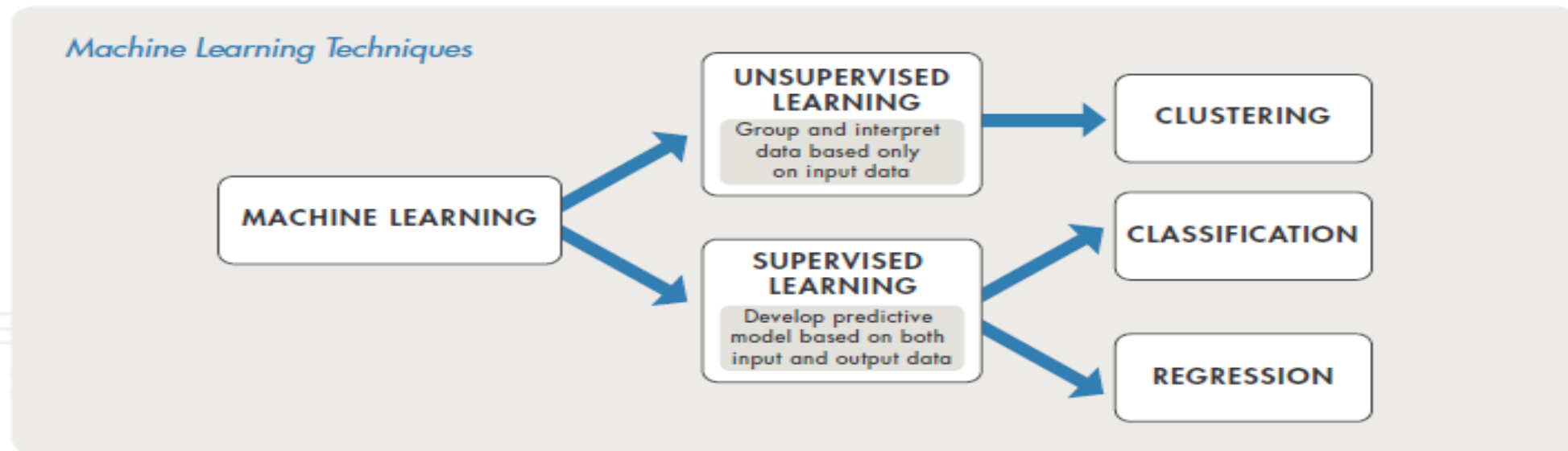
- Every machine learning workflow begins with three questions:
  - What kind of data are you working with?
  - What insights do you want to get from it?
  - How and where will those insights be applied?





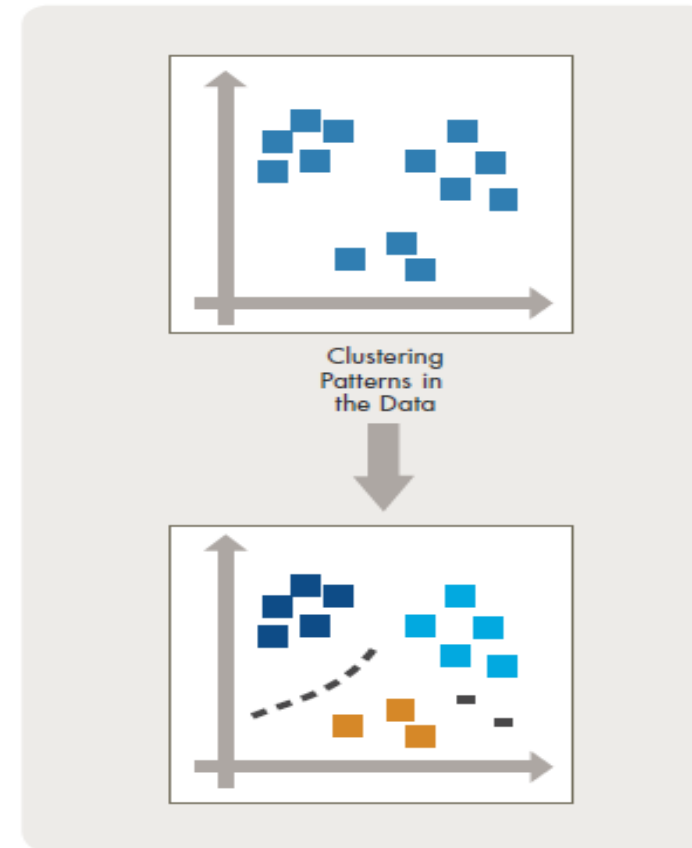


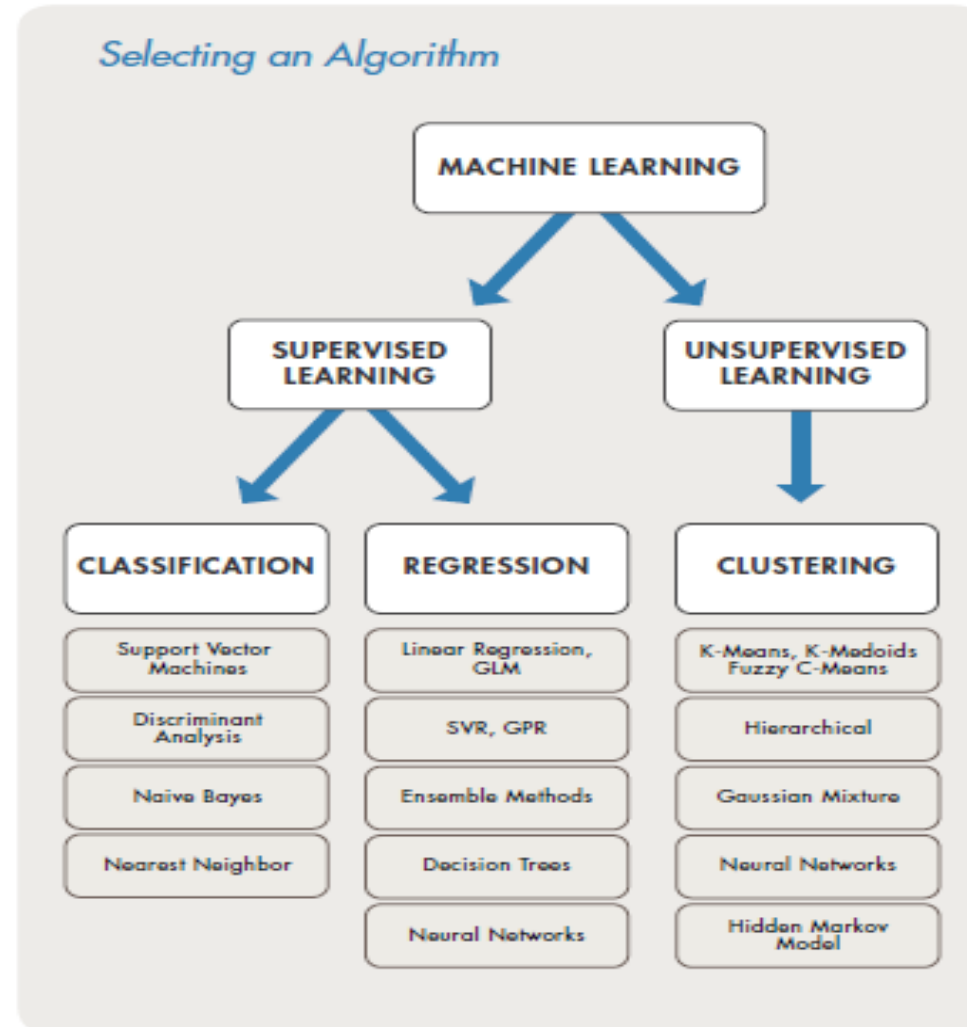
- Machine learning uses two types of techniques:
  - Supervised learning: trains a model on known input and output data so that it can predict future outputs
  - Unsupervised learning: finds hidden patterns or intrinsic structures in input data.



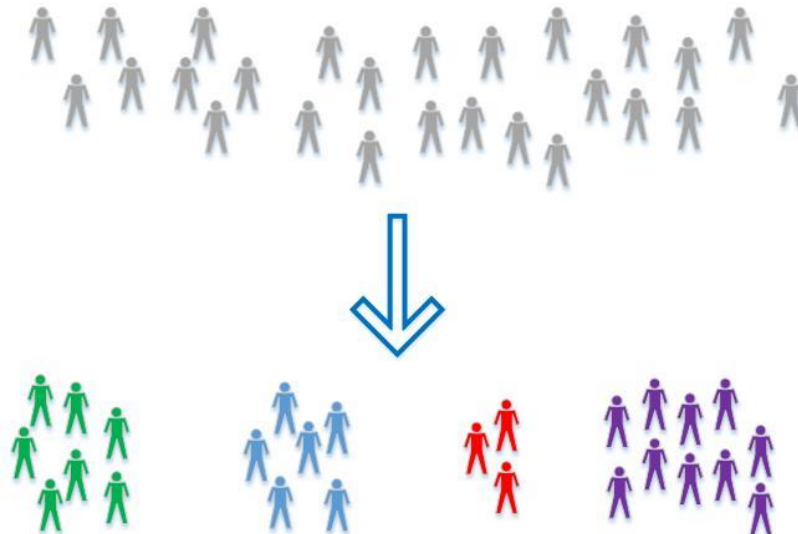
- Makes predictions based on evidence in the presence of uncertainty.
- Takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data.
- Uses classification and regression techniques to develop predictive models.
- Classification techniques
  - predict discrete responses—for example, whether an email is genuine or spam, or whether a tumor is cancerous or benign.
  - classify input data into categories. Typical applications include medical imaging, speech recognition, and credit scoring.
- Regression techniques
  - predict continuous responses— for example, changes in temperature or fluctuations in power demand.
  - Typical applications include electricity load forecasting and algorithmic trading.

- Finds hidden patterns or intrinsic structures in data. It is used to draw inferences from datasets consisting of input data without labeled responses.
- Clustering
  - the most common unsupervised learning technique.
  - used for exploratory data analysis to find hidden patterns or groupings in data.
  - Applications for clustering include gene sequence analysis, market research, and object recognition.

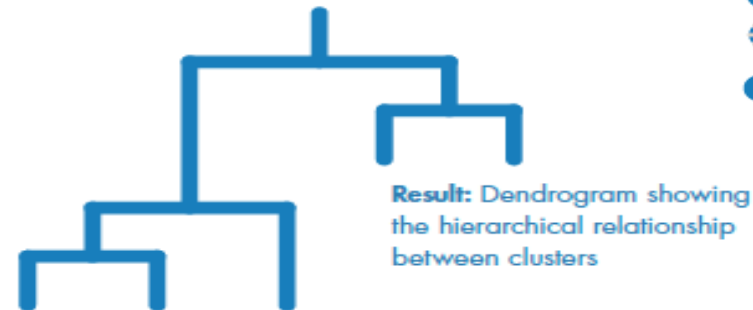
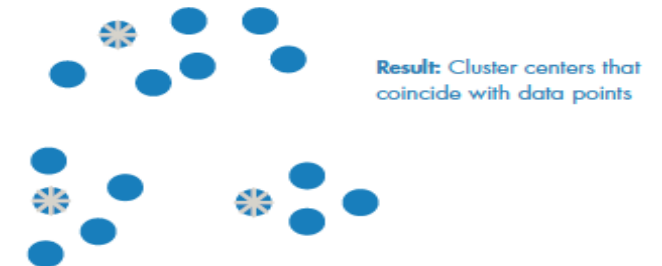




- Most unsupervised learning techniques are a form of cluster analysis.
- Cluster analysis
  - data is partitioned into groups based on some measure of similarity or shared characteristic.
  - Clusters are formed so that objects in the same cluster are very similar and objects in different clusters are very distinct.
- Clustering algorithms fall into two broad groups:
  - Hard clustering, where each data point belongs to only one cluster
  - Soft clustering, where each data point can belong to more than one cluster
- You can use hard or soft clustering techniques if you already know the possible data groupings.



- k-Means
- k-Medoids
- Hierarchical Clustering
- Self-Organizing Map



- **Fuzzy c-Means**



**Result:** Cluster centers (similar to k-means) but with fuzziness so that points may belong to more than one cluster

- **Gaussian Mixture Model**



**Result:** A model of Gaussian distributions that give probabilities of a point being in a cluster

- All supervised learning techniques are a form of classification or regression.
- Classification techniques predict discrete
- Regression techniques predict continuous



- **Logistic Regression**
- **k Nearest Neighbor (kNN)**
- **Support Vector Machine (SVM)**
- **Neural Network**
- **Naïve Bayes**
- **Discriminant Analysis**
- **Decision Tree**

- **Linear Regression**
- **Nonlinear Regression**
- **Gaussian Process Regression Model**
- **SVM Regression**
- **Regression Tree**

- Team Mitchell is using ML in a lot of applications:
  - Smart Recommender: A recommendation engine for car insurance claims on broken parts
  - Image Processing: to recognize car's color from the image.
  - Image Processing: identify point of impact on car based on damaged area
  - Driver injury detection
  - Smart triach: a questionnaire based insurance recommender.

- Learn more about Machine Learning
  - Coursera Courses
  - CDP: Infogain Certified Machine Learning Developer
- Skills Required:
  - Data Analytics basic knowledge
  - Programming language working knowledge: Python/R/Java
- Time Commitment:
  - 20 weeks (5-6 hrs/week) for a good hold





### **Infogain Corporation, HQ**

485 Alberto Way Los Gatos,  
CA 95032 USA  
Phone: 408-355-6000  
Fax: 408-355-7000

### **Pune**

7th Floor, Bhalerao Towers, CTS No.1669 -  
1670, Behind Hotel Pride,  
Shivaji Nagar, Pune - 411005  
Phone : +91-20-66236700

### **Infogain Irvine**

41 Corporate Park,  
Suite 390 Irvine, CA 2606 USA  
Phone: 949-223-5100  
Fax: 949-223-5110

### **Infogain Austin**

Stratum Executive Center Building D  
11044 Research Boulevard Suite 200  
Austin, Texas 78759

### **Noida**

A-16, Sector 60, Noida Gautam Budh agar,  
201301 (U.P.) India  
Phone: +91-120-2445144  
Fax: +91-120-2580406

### **Dubai**

P O Box 500588 Office No.105,  
Building No. 4, Dubai Outsource Zone,  
Dubai, United Arab Emirates  
Tel: +971-4-458-7336