Instructions:

=> If course materials brought you here, scroll or use links to sections.
=> Use links on downloaded.pdf! In git, download arrow is on the right above doc visual.

Titanic Data	PreProcess + Supervised & Unsupervised		

M.2.Assignment - Titanic Data

M	Topic & Assignment

M2

Titanic data mining analysis

A. **Background and overviews**

- https://www.rdocumentation.org/packages/titanic/versions/0.1.0
- https://www.kaggle.com/competitions/titanic/overview
- https://www.encyclopedia-titanica.org/

The Titanic DataFrames describe the survival status of individual Titanic passengers, not the crew, with ages for ~half the passengers. One of the original sources is Eaton & Haas (1994) Titanic: Triumph and Tragedy, Patrick Stephens Ltd includes a passenger list created by many researchers and edited by Michael A. Findlay [1].

B. **Interesting models - built in R code for display convenience**

```
> data <- read.csv('titanic.csv')
```

- # Linear regression model
- model <- lm(survived ~ age + sex + pclass + sibsp + parch, data = data)

- **Binomial** Predicting survival based on age, sex, and passenger class
- model <- glm(survived ~ age + sex + pclass, data = titanic, family = binomial)

- **Poisson** - Predicting the count of siblings/spouses based on passenger age
- model <- glm(sibsp ~ age, data = titanic, family = poisson) summary(model)

- **Neg.Binomial** - Predict count of parents/children by passenger age and sex
- model <- glm.nb(parch ~ age + sex, data = titanic) summary(model)

C. **Data <class.github>**

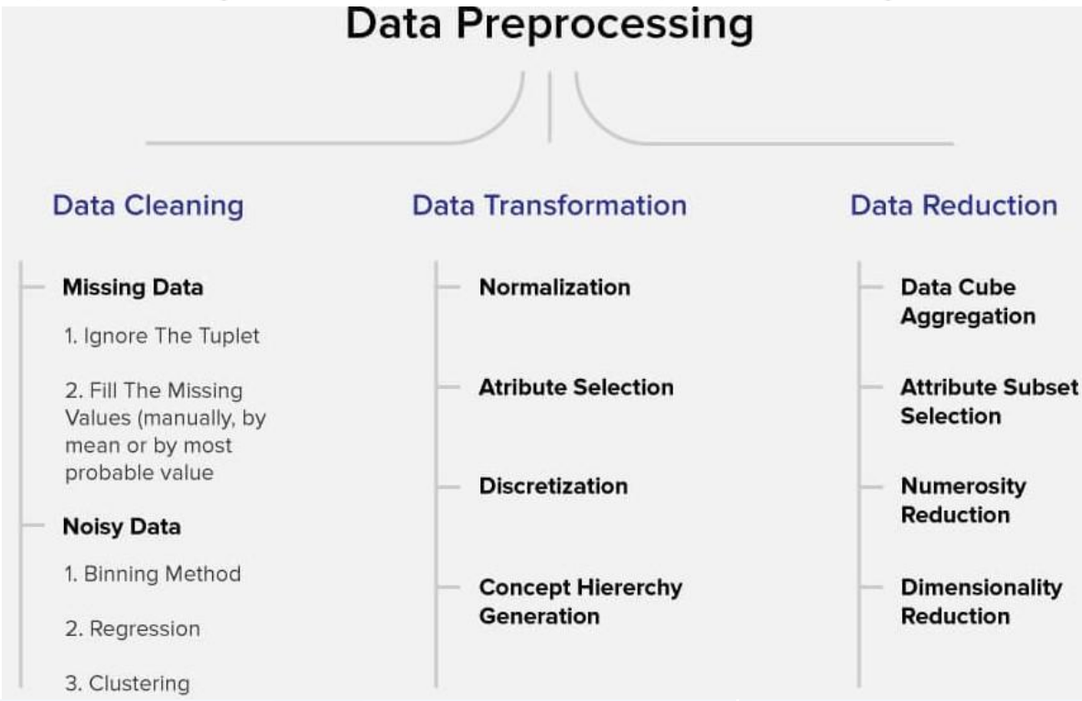
- raw data; unsplit and preprocessed [source: https://hbiostat.org/data/ <titanic.3>
- train, test; from kaggle

D. **Data dictionary**

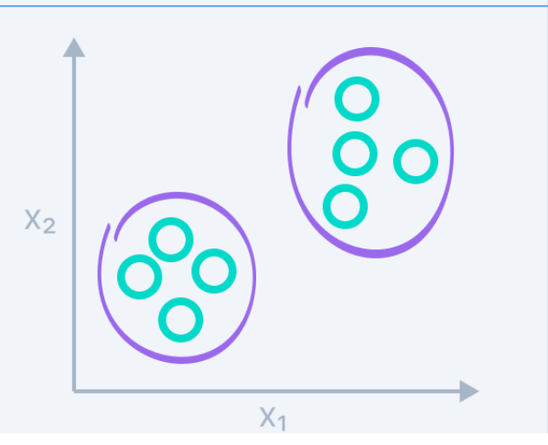
passengerid	sequential unique id
survived	0=no, 1=yes
pclass	1,2,3:passenger class (1st, 2nd, 3rd); proxy for socio-economic class
name	Christian name
sex	male, female
age	00, NA, blank. in years; some infants w fractional values
sibsp	number of siblings and spouses aboard
parch	<parent.child> #parents or chil
ticket	alpha, numeric, character
fare	0.0000 decimals
cabin	C#, blank,
embarked	C, Q, S <Cherbourg, Southampton, and Queenstown>

References:

1. Harrell Jr, F.E.,(2002). Titanic data, Vanderbuilt biostatistics datasets. Vanderbilt University. Retrieved from: https://hbiostat.org/data/repo/titanic.html. Retrieved on 05.15.2023.



Unsupervised learning	Supervised learning
Input data is unlabeled	Input data is labeled
Has no feedback mechanism	Has a feedback mechanism
Assigns properties of given data to classify it	Data is classified based on the training dataset
Divided into Clustering & Association	Divided into Regression & Classification
Used for analysis	Used for prediction
Algorithms include: k-means clustering, hierarchical clustering, apriori algorithm	Algorithms include: decision trees, logistic regressions, support vector machine
A unknown number of classes	A known number of classes



Dimensionality Reduction

- Feature Elicitation
- Meaningful Compression
- Structure Discovery
- Big data visualization

Clustering

- Recommender Systems
- Targeted Marketing
- Customer Segmentation

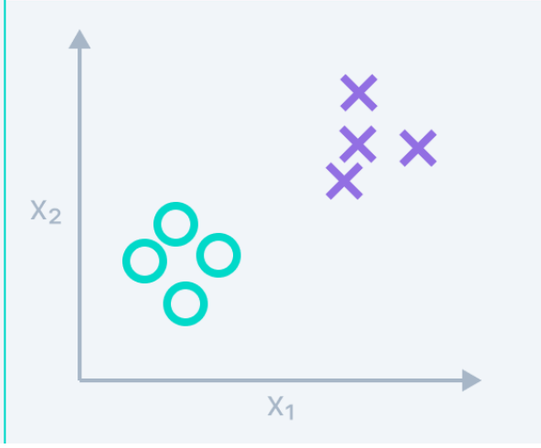
Classification

- Identity Fraud Detection
- Image Classification
- Customer Retention
- Diagnostics

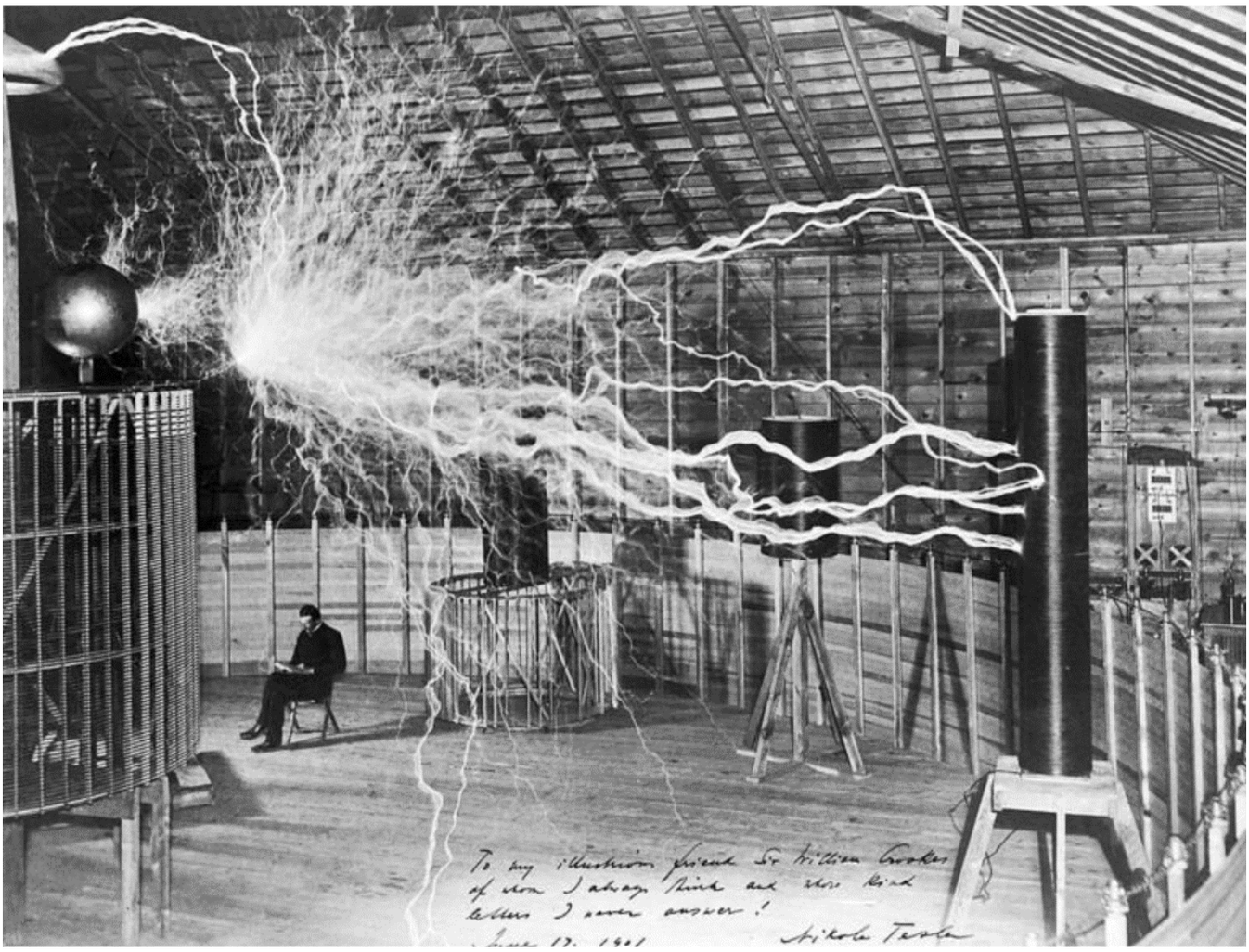
Regression

- Population Growth Predictio
- Estimating life expectancy
- Market Forecasting
- Weather Forecasting
- Advertising Popularity Predi

- Real-time decisions
- Game AI
- Robot Navigation
- Learning Tasks
- Skill Acquisition



retrieved from: https://www.v7labs.com/blog/supervised-vs-unsupervised-learning


To my illustrious friend Sir William Crookes
of whom I always think and who kind
letters I never answer!
Nikola Tesla

Master templates

my.header.2

Mod	Topic & Assignment
2	

my.header.1

Wk	Focus & Medium	Weekly Topic & Assignment
0		

Wk	Weekly Topic & Assignment
11 Mar	**Header:** **1) What is** 1) Perhaps one of th **2) What isn't** a) Getting your n **3) The mechanics and process** ● Orient **4) problem** **Templated techniques** help you quickly

Wk	Weekly Topic & Assignment
11 Mar	**Templated writing techniques** hel **Use kernel sentences:** simple, declarative, active sentences (N.Chomsky) Use of clear and concise language that is free of jargon and technical terms focuses the reader. a) Joh **1. Template: how.to. abc** 1.1. item.1: Tai 1.2. item.s: U 1.3. item.: Us 1.4. item.: Ack **1.5.** item.5: abdc **Scenario**: The