



# **Introduction to Data Science**

Week 1

# Welcome to Data Science

---

Introduction to theoretical background and practical aspect of data science

Connection to other fields including Statistics, Linear algebra  
Machine Learning

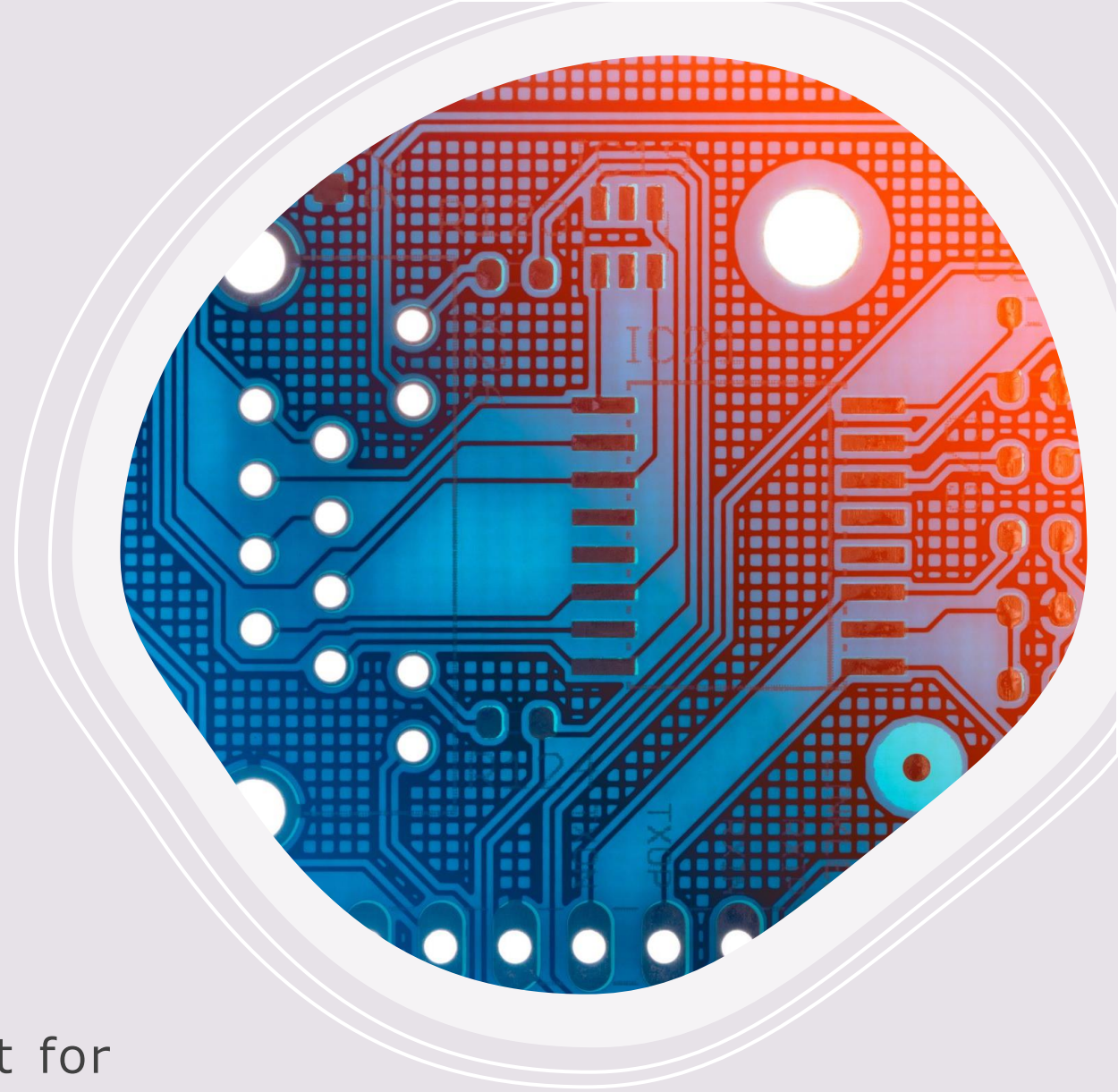
Number of real-world problems will be discussed

Module uses Python programming language and Jupiter environment

# Topic 1

Introduction to data science:

- Scope of data science
- Impact in academia and industry
- Main concepts in data science will be discussed
- Introduction to Python programming language
- Introduction to the Jupiter environment for Python development



# Topic 2

Introduction to data processing:

- Various data processing techniques
- Introduction to Python libraries like NumPy and pandas
- Understanding measures of central tendency and measures of spread
- Concepts of Linear Algebra





# Topic 3

Introduction to data visualization:

- Understanding data visualization
- Different approaches to handling qualitative and quantitative data will be reviewed
- Different types of diagrams and their impact on the intended audience will be discussed and demonstrated
- Real world examples using Python library Matplotlib



# Topic 4

## Introduction to statistics

- Theoretical foundations of statistics
- Main data types in context of statistics
- Descriptive and inferential statistics
- Processing of different types of variables





# Topic 5

Introduction to machine learning using scikit learn

- Theoretical foundations of machine learning
- Main Python libraries for machine learning
- Concept of model validation
- Techniques for selecting the best model
- Data processing using feature engineering

# Topic 6

## Basic Text Processing

- Principles of processing unstructured text
- Concept of Regular expression
- Real world application





# Topic 7

## Introduction to Natural Language Processing

- Python libraries for processing natural data
- Techniques for representing word meanings
- Practical applications of NLP
- NLP pipelines



The background is a complex, abstract composition. It features a network of white lines connecting various points, creating a web-like structure. Overlaid on this are several geometric shapes, including triangles and polygons, in shades of blue and purple. There are also elements resembling data visualizations: a line graph with axes in the top left, a heatmap with a color gradient in the bottom right, and several instances of binary code (0s and 1s) scattered throughout. The overall color palette is dominated by deep blues, purples, and a hint of pink/purple on the right side where the text is located.

# Topic 8

## Introduction of Data Visualization

- Advanced data visualization
- Large and multidimensional datasets
- Understanding heatmaps and parallel coordinates

# Topic 9

---

## Machine Learning :

- Advanced techniques in Machine Learning
- Role of Bayes' Theorem in Supervised Learning
- Practical examples of Linear Regression, Support Vector Machines and Decision Trees
- Unsupervised Learning using KMeans Clustering





# Topic 10

Case Studies:

Real world examples