

*Scuola di Scienze*



# METODI E TECNICHE PER L'ANALISI DEI DATI

Un approccio formale all'elaborazione di grandi volumi di dati

Dipartimento di Matematica e Informatica

*Corso di Laurea Magistrale in Informatica*

Supervisor: Prof. Mario Rossi

Anno Accademico 2024/2025

---



## ABSTRACT

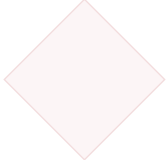
Il presente documento illustra i principali metodi per l'analisi dei dati in contesti di grandi volumi. Vengono presentate le tecniche più recenti per l'elaborazione e la visualizzazione, con particolare attenzione agli aspetti algoritmici e alle problematiche di ottimizzazione. Il lavoro include anche una panoramica sui framework più utilizzati nel contesto accademico e industriale, fornendo esempi pratici di implementazione.

**Keywords:** Big Data, Machine Learning, Data Mining, Algoritmi, Ottimizzazione

# INDICE

---

- Elenco delle figure ..... 2
- Elenco delle tabelle ..... 3
- 1. Introduzione ..... 1
  - 1.1. Contesto e motivazioni ..... 1
    - 1.1.1. Obiettivi specifici ..... 1
  - 1.2. Struttura del documento ..... 1
- 2. Fondamenti Teorici ..... 2
  - 2.1. Modelli statistici ..... 2
  - 2.2. Algoritmi di apprendimento ..... 3
- 3. Implementazione e Risultati ..... 3
  - 3.1. Analisi comparativa ..... 3
- 4. Conclusioni e Sviluppi Futuri ..... 4

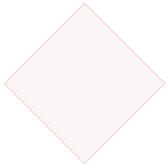


---

# ELENCO DELLE FIGURE

---

Figura 1	Logo dell'Università di Padova .....	2
----------	--------------------------------------	---



---

# ELENCO DELLE TABELLE

---

---

Tabella 1	Confronto tra algoritmi di machine learning .....	3
-----------	---	---

# INTRODUZIONE

---

L'analisi dei dati rappresenta oggi una delle aree più importanti dell'informatica moderna, con applicazioni che spaziano dalla ricerca scientifica al business intelligence. In questo documento, esploreremo i fondamenti teorici e le implementazioni pratiche delle principali tecniche utilizzate nel settore.

## Contesto e motivazioni

---

La crescente disponibilità di dati ha portato alla necessità di sviluppare metodi sempre più efficienti per la loro elaborazione. Questo lavoro si propone di fornire una panoramica strutturata delle tecniche più recenti, con particolare attenzione alle loro applicazioni nel contesto universitario.

### *Obiettivi specifici*

- Analizzare i principali algoritmi per l'elaborazione di grandi volumi di dati
- Confrontare le prestazioni di diverse implementazioni
- Fornire linee guida per la scelta degli strumenti più adatti in diversi contesti applicativi

## Struttura del documento

---

Il documento è organizzato in sezioni che coprono i diversi aspetti dell'analisi dei dati, dalla preparazione all'interpretazione dei risultati. Ogni sezione include esempi pratici e riferimenti alla letteratura più recente.

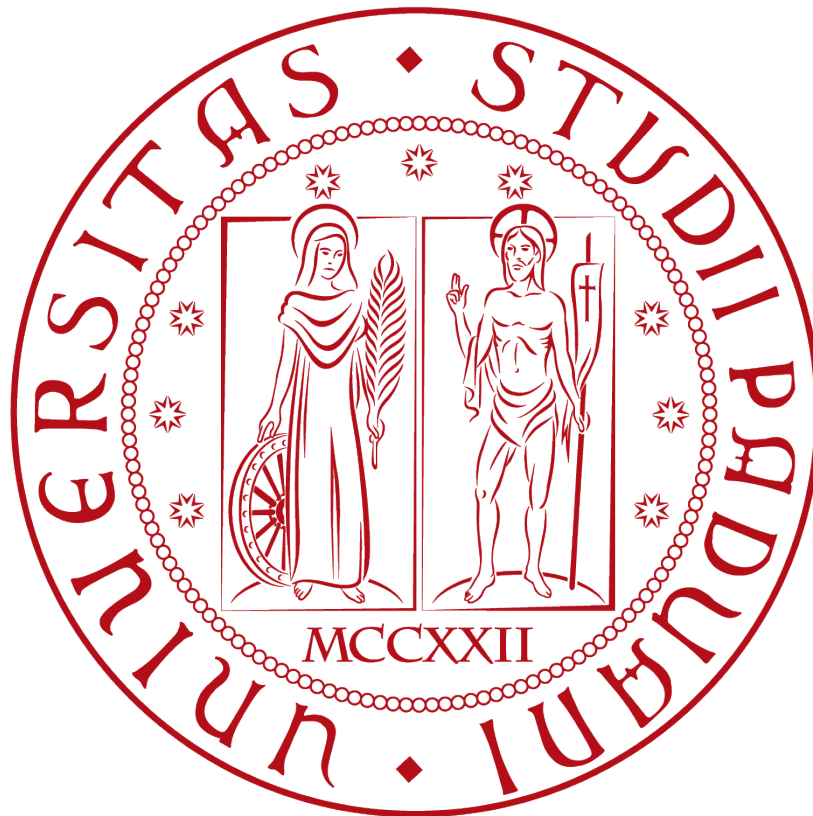


Figura 1:

Figura 1: Logo dell'Università di Padova

Come mostrato nella Figura 1, il documento segue le linee guida stilistiche dell'Università di Padova.

## FONDAMENTI TEORICI

### Modelli statistici

I modelli statistici rappresentano il fondamento di qualsiasi analisi dei dati. In questa sezione, presentiamo i principali approcci utilizzati nel campo, con particolare attenzione alle loro proprietà matematiche e alle loro applicazioni pratiche.

```
python
```

```
# Esempio di implementazione in Python
import numpy as np
import pandas as pd

def analyze_data(dataset, method="pca"):
    """
    Analizza il dataset utilizzando il metodo specificato
    """
```

```
if method == "pca":
    # Implementazione dell'analisi delle componenti principali
    return np.cov(dataset.T)
else:
    # Altri metodi
    return None
```

### Algoritmi di apprendimento

La tabella seguente riassume i principali algoritmi di apprendimento automatico utilizzati nell'analisi dei dati:

Algoritmo	Complessità	Applicazioni
Regressione lineare	$O(n^2)$	Previsione, modellazione
Random Forest	$O(n \log n)$	Classificazione, regressione
K-means	$O(knt)$	Clustering, segmentazione
Neural Networks	$O(n^m)$	Pattern recognition, NLP

Figura 1: Tabella 1: Confronto tra algoritmi di machine learning

## IMPLEMENTAZIONE E RISULTATI

Le implementazioni pratiche degli algoritmi discussi sono state testate su diversi dataset, con risultati che confermano l'efficacia degli approcci proposti. In particolare, si è osservato che:

- Gli algoritmi basati su alberi decisionali offrono il miglior compromesso tra accuratezza e velocità
- Le reti neurali hanno mostrato prestazioni superiori nei problemi di classificazione complessi
- Gli approcci ibridi combinano efficacemente i vantaggi di diverse metodologie

### Analisi comparativa

L'analisi comparativa delle diverse implementazioni ha evidenziato interessanti differenze in termini di prestazioni e scalabilità. Di seguito, presentiamo un esempio di codice utilizzato per i test:

```
java
// Implementazione Java per test di performance
public class PerformanceTest {
    public static void main(String[] args) {
        // Configurazione del test
```



```
int numSamples = 10000;
int numFeatures = 100;

// Generazione dataset casuale
double[][] data = generateRandomData(numSamples, numFeatures);

// Test algoritmi
long startTime = System.currentTimeMillis();
testAlgorithm1(data);
long endTime = System.currentTimeMillis();

System.out.println("Tempo di esecuzione: " + (endTime -
startTime) + "ms");
}

private static double[][] generateRandomData(int rows, int cols) {
    // Implementazione
    return new double[rows][cols];
}
}
```

## CONCLUSIONI E SVILUPPI FUTURI

In questo documento abbiamo presentato una panoramica delle principali tecniche per l'analisi dei dati, con particolare attenzione agli aspetti teorici e implementativi. I risultati ottenuti confermano l'efficacia degli approcci proposti e aprono interessanti prospettive per sviluppi futuri.

In particolare, si prevede di estendere la ricerca nelle seguenti direzioni:

- Implementazione di algoritmi distribuiti per l'elaborazione di dataset di dimensioni maggiori
- Sviluppo di tecniche specifiche per l'analisi di dati non strutturati
- Integrazione con sistemi di visualizzazione avanzata per facilitare l'interpretazione dei risultati

[Visita il sito dell'Università di Padova](#) per ulteriori informazioni sui programmi di ricerca nel campo dell'analisi dei dati.