



# MaDS: Matematica per Data Science



## GUIDA AI PIANI DI STUDIO 2022-2023

Storicamente la matematica ha dato, e dà tuttora, contributi fondamentali al problema della trasmissione sicura dell'informazione attraverso la **crittografia**, che ha le sue basi nella teoria dei numeri. Inoltre i matematici hanno sviluppato tecniche e algoritmi per l'**analisi** e il **trattamento di dati**, che hanno applicato alle ricerche delle parole chiave in un testo o di oggetti in **cataloghi multimediali**, al **trattamento delle immagini**, all'analisi del **flusso di dati sulle reti**, all'analisi di **sequenze dati naturali**, per esempio biometrici, meteorologici o sismici. Negli ultimi anni l'esplosione delle tecniche di **Deep Learning** nelle applicazioni ha aggiunto nuovi strumenti, e sembra suggerire che, per ottenere risposte su un fenomeno in esame, basti far leggere i relativi dati a opportuni **algoritmi intelligenti**. Tuttavia, risposte ottenute senza un reale controllo su queste black-box non permettono di valutarne l'affidabilità, che è invece indispensabile per un vero progresso delle conoscenze. L'**affidabilità** è inoltre cruciale nelle applicazioni in contesti in cui le risposte contribuiscono a decisioni di grande impatto sociale, come è il caso di diagnosi e terapie mediche. Per questo è necessario capire come funzionano questi algoritmi, quali elementi del fenomeno prendono in considerazione, e come li elaborano. È altrettanto importante confrontare pregi e limiti di questi strumenti in relazione alla modellistica matematica, perché l'interazione tra i due approcci permette di migliorare conoscenze e qualità delle applicazioni.

### OBIETTIVI

Laureate e laureati con una solida e avanzata preparazione in matematica possono avere un ruolo importante in queste sfide, sono in grado di **comprendere i fondamentali** su cui si basano queste nuove tecniche e conoscono le **potenzialità esplicative e predittive** degli strumenti matematici coinvolti.

L'obiettivo del curriculum è formare matematiche e matematici con ampie competenze nei settori dell'**algebra**, della **probabilità**, della **statistica** della **fisica matematica**, dell'**analisi numerica**, della **ricerca operativa**, che siano in grado di affrontare e risolvere i problemi posti nell'ambito della **Data Science**. Svilupperemo negli anni questo nuovo curriculum, anche approfondendo l'interazione con i corsi di studio che offrono insegnamenti più specializzati. Per questo motivo incoraggiamo l'inserimento nel piano di studi anche di corsi esterni in linea con queste applicazioni.

## CORSI OBBLIGATORI

### I anno - I semestre

- **Istituzioni di probabilità** (9 CFU)
- **Istituzioni di analisi numerica** (9 CFU)
- **Istituzioni di algebra e geometria** (9 CFU)
- **Statistica matematica** (6 CFU)

### II anno – I semestre

- **Programmazione matematica** (9 CFU)  
(Introduzione all'ottimizzazione 6 CFU, Metodi per Big Data e Machine Learning 3 CFU)
- Inglese scientifico (4 CFU, idoneità)
- Altre conoscenza utili per l'inserimento nel mondo del lavoro (3 CFU, idoneità)

## CORSI OPZIONALI SPECIFICI DELL'INDIRIZZO

Nei primi tre semestri dei due anni vanno scelti **6 corsi opzionali** da 6 CFU nei seguenti gruppi (si noti che i corsi contrassegnati con **ENG** sono impartiti in lingua inglese)

- un corso **opzionale applicato**  
**Analisi di sequenze di dati**  
**Data mining**  
**Mathematical Models for Neural Networks** **ENG**  
**Elementi di probabilità e statistica per Data Science**
- un corso **opzionale algebrico**  
**Combinatorics** **ENG**                      **Matematica discreta**
- un corso **opzionale informatico**  
**Teoria degli algoritmi**                      **Cryptography** **ENG**
- due corsi **integrativi** tra  
**Fourier Analysis** **ENG**                      **Analisi di sequenze di dati**  
**Data mining**                                      **Mathematical Models for Neural Networks** **ENG**  
**Combinatorics** **ENG**                              **Matematica discreta**  
**Teoria degli algoritmi**                              **Cryptography** **ENG**  
**Processi stocastici**                              **Computational Mathematics** **ENG**  
**Metodi matematici in meccanica statistica**  
**Elementi di probabilità e statistica per Data Science**



Il curriculum MaDS fornisce le basi per la formazione matematica avanzata attraverso i corsi di **Istituzioni** da 9 CFU di **algebra e geometria**, **probabilità**, **analisi numerica**, **programmazione matematica** (ricerca operativa), e con il corso di **statistica matematica** da 6 CFU.

L'ossatura della formazione matematica nella direzione della **data science**, è formata da corsi di **algebra**, **probabilità**, **fisica matematica**, **analisi numerica**, tra cui si può scegliere a seconda dei propri interessi.

**Analisi di sequenze dati**: è un corso di strumenti statistici, probabilistici, informatici, per lo studio di modelli di serie temporali di dati (per esempio dati biometrici, geologici, o generati da attività umane come le sequenze di simboli di un testo).

**Combinatorics (ENG)**: il corso si propone di introdurre i concetti fondamentali di teoria dei grafi, sempre più utilizzata in vaste aree applicative, dalle reti mobili, al traffico di dati, al web: grafi semplici, planarità, invarianti di grafi, grafi orientati, teoria di Ramsey, cenni sui grafi random.

**Data mining**: il corso illustra alcune tecniche di algebra lineare numerica, molto utilizzate nell'estrazione di informazioni utili da grandi quantità di dati (data mining). Verranno affrontati alcuni problemi importanti nelle applicazioni: compressione, classificazione, clustering. Per risolverli è necessario affrontare in spazi di dimensione enorme la risoluzione numerica di sistemi lineari, i problemi agli autovalori e valori singolari, il calcolo di funzioni di matrici e la gestione di grafi.

**Elementi di probabilità e statistica per data science**: nei corsi introduttivi alla probabilità e alla statistica, si considerano grandi quantità di dati in spazi di poche dimensioni (una o due). Questo corso introduce ai modelli e alle tecniche matematiche per lo studio di dati in spazi di alta dimensione, che è una delle caratteristiche peculiari dei "Big Data".

**Matematica discreta**: combinatoria enumerativa (principio di Inclusione-esclusione, le serie generatrici, i "rook polynomials"); teoria degli insiemi parzialmente ordinati e dei reticoli; combinatoria delle permutazioni; statistiche di permutazioni; permutazioni random.

**Mathematical Models for Neural Networks (ENG)**: introduzione allo studio delle reti neurali artificiali, in particolare reti ad attrattori (utilizzate come modelli di memoria associativa) e reti multistrato (utilizzate nell'apprendimento automatico) e agli algoritmi utilizzati per l'addestramento e il funzionamento di queste reti. Strumenti matematici per lo studio e il controllo di questi sistemi, derivanti prevalentemente dalla meccanica statistica, dalla teoria della probabilità e dalla statistica matematica.

Inoltre sarà indispensabile conoscere l'approccio informatico, e per questo sono offerti i corsi di **Teoria degli algoritmi**, incentrato sugli aspetti matematici e fondazionali del Machine Learning, e **Cryptography (ENG)**, che introduce alla crittografia.



Questo è lo schema delle scelte distribuite nei due semestri. Si noti che, in linea generale, i corsi del I anno richiedono meno prerequisiti dei corsi del II anno, che sono più specialistici e avviano alla tesi. Nella formazione del piano di studi si raccomanda di considerare i prerequisiti indicati dai docenti dei singoli corsi che si intende seguire.

Esami	I anno I semestre	I anno II semestre	II anno I semestre
Obbligatori da 9	<b>Istituzioni di analisi numerica</b> <b>istituzioni di probabilità</b> <b>Ist. di algebra e geometria</b>	<b>Programmazione matematica</b>	
Obbligatori da 6	<b>Statistica matematica</b>		
1 nel gruppo applicato		Analisi di sequenze dati Data mining	Mathematical Models for Neural Networks Elementi di probabilità e statistica per data science
1 nel gruppo algebrico		Matematica discreta	Combinatorics
1 nel gruppo informatico		Teoria degli algoritmi	Cryptography
2 nel gruppo integrativo		Analisi di sequenze dati Data mining Computational Mathematics Matematica discreta Teoria degli algoritmi Metodi matematici in meccanica statistica Processi stocastici	Mathematical Models for Neural Networks Elementi di probabilità e statistica per data science Combinatorics Cryptography Fourier Analysis

## CORSI LIBERI A SCELTA

Come da regolamento didattico di Ateneo, si possono scegliere liberamente due corsi da 6 CFU, sia tra quelli offerti dai corsi di studi in Matematica Applicata e in Matematica, sia tra quelli offerti dai altri corsi di studio magistrali e triennali della Sapienza. Uno va inserito al II semestre del I anno, e uno al I semestre del II anno.

Durante i seminari MoMA <http://www1.mat.uniroma1.it/ricerca/seminari/moma/> di questi anni, abbiamo preso contatto con vari colleghi di altre discipline interessati alla Data Science. Cogliamo quindi l'occasione per segnalare qui dei corsi da 6 CFU da loro tenuti, insieme ad altri corsi selezionati tra quelli offerti dalla Sapienza, che posso certamente essere fruiti da studenti delle magistrali in matematica interessanti alla Data Science.

Sergio Barbarossa: Signal Processing for Big Data (**ENG** ING-INF03) Laura magistrale in Data Science

Roberto Navigili: Natural Language Processing (**ENG** INF01) Laura magistrale in Computer Science

docenti vari: Fisica dei sistemi complessi (FIS03) Laurea magistrale in Fisica

Pierpaolo Brutti: Statistical Learning (**ENG**SECS-S/01) Laurea magistrale in Data Science

Emanuele Rodolà: Deep learnig and applied artificial intelligence (**ENG** NF/01) Laurea magistrale in Computer Sciece – Informatica

Lorenzo Carlucci: Mathematical logic for computer science (**ENG** INF/01) Laurea magistrale in Computer Sciee – Informatica

Tiziana Calamoneri: Network alghoritms (**ENG** INF/01) Laurea magistrale in Computer Sciee – Informatica

Gaetano Scarano: Elaborazione numerica dei segnali (ING-INF/03) Laurea triennale in Ingegneria delle comunicazioni

Stefania Colonnese: Elaborazione delle immagini (ING-INF/03) Laurea magistrale in Ingegneria delle comunicazioni

## ESEMPI DI PIANI DI STUDIO

Il contributo trasversale delle varie discipline matematiche all'interno di questo curriculum può essere organizzato in 3 principali direzioni, per le quali presentiamo delle tabelle con una selezione ragionata delle opzioni per i corsi a scelta. Ricordiamo però che potete costruire il vostro piano di studi seguendo la tabella generale, secondo i vostri interessi e le vostre inclinazioni.

In magenta sono evidenziati i **corsi obbligatori**,  
in arancione i **corsi caratterizzanti**,  
in blu i **corsi tra cui scegliere**.

### Modelli matematici per Data Science

Questo percorso fornisce una solida base matematica per lo studio dei recenti modelli usati per l'analisi dei dati, in particolare per i dati in alta dimensione.

Esami a scelta	I anno – I semestre	I anno - II semestre	II anno - I semestre
Obbligatori da 9	<b>Istituzioni di analisi numerica</b> <b>Istituzioni di probabilità</b> <b>Ist. di algebra e geometria</b>	<b>Programmazione matematica</b>	
Obbligatori da 6	<b>Statistica matematica</b>		
1 nel gruppo applicato			<b>Mathematical Models for Neural Networks</b>
1 nel gruppo algebrico		<b>Matematica discreta</b>	<b>Combinatorics</b>
1 nel gruppo informatico		<b>Teoria degli algoritmi</b>	
2 nel gruppo integrativo		<b>Analisi di sequenze dati</b> <b>Processi stocastici</b> <b>Metodi matematici in meccanica statistica</b> <b>Data mining</b>	<b>Elementi di probabilità e statistica per data science</b> <b>Fourier Analysis</b>



### Matematica discreta e crittografia

Questo percorso presenta i metodi di modellizzazione discreta, algebrico/geometrici e numerico/informatici, che trovano applicazione nello studio di problemi su reti e nella sicurezza informatica.

Esami a scelta	I anno – I semestre	I anno - II semestre	II anno - I semestre
Obbligatori da 9	<b>Istituzioni di analisi numerica istituzioni di probabilità Ist. di algebra e geometria</b>	<b>Programmazione matematica</b>	
Obbligatori da 6	<b>Statistica matematica</b>		
1 nel gruppo applicato		<b>Data mining</b>	
1 nel gruppo algebrico			<b>Combinatorics</b>
1 nel gruppo informatico			<b>Cryptography</b>
2 nel gruppo integrativo		<b>Matematica discreta Analisi di sequenze dati Teoria degli algoritmi Processi stocastici</b>	

### Analisi e trattamento dei dati

Questo percorso si concentra sulle principali tecniche matematiche per trattare dati, utili per in molte applicazioni.

Esami a scelta	I anno – I semestre	I anno - II semestre	II anno - I semestre
Obbligatori da 9	<b>Istituzioni di analisi numerica istituzioni di probabilità Ist. di algebra e geometria</b>	<b>Programmazione matematica</b>	
Obbligatori da 6	<b>Statistica matematica</b>		
1 nel gruppo applicato		<b>Analisi di sequenze dati</b>	
1 nel gruppo algebrico		<b>Matematica discreta</b>	<b>Combinatorics</b>
1 nel gruppo informatico		<b>Teoria degli algoritmi</b>	<b>Cryptography</b>
2 nel gruppo integrativo		<b>Data mining Processi stocastici</b>	<b>Fourier Analysis</b>