

# Machine learning per il cambiamento climatico: applicazione a studi di abbondanza dei cetacei

Tesi di Laurea in Sistemi Multimediali

Relatore:

Prof. **Giovanni Dimauro**

Correlatore:

Dott.ssa **Rosalia Maglietta**

Laureando:

**Leonardo Saccotelli**



# Sommario

---

- Introduzione
- Descrizione del data set
- Analisi statistica dei dati
- Analisi dei modelli di regressione sviluppati
- Discussione dei risultati
- Conclusioni e sviluppi futuri



# Introduzione

---

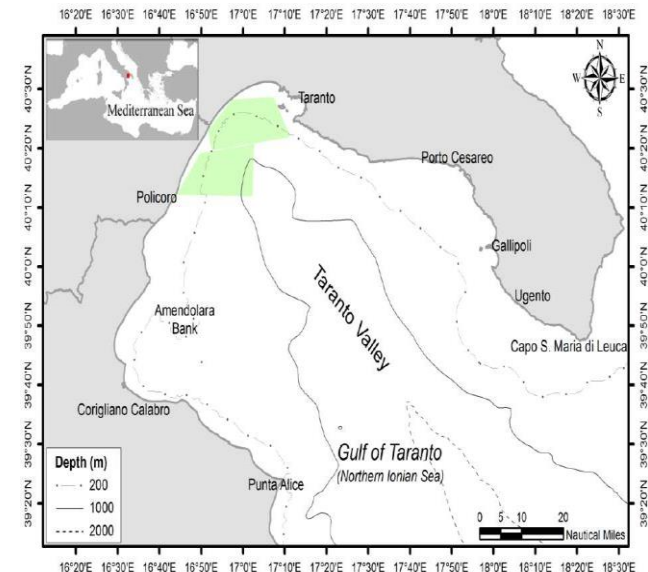
Lo studio dei cetacei è fondamentale per comprendere come gli ecosistemi marini stanno mutando e quali sono gli effetti principali di questi cambiamenti.

Identificare e attuare misure di conservazione che favoriscano la resistenza dei cetacei ai cambiamenti indotti dal clima è doveroso poiché tali cambiamenti possono potenzialmente influenzare negativamente lo stato di conservazione e la distribuzione dei cetacei.

# Introduzione

Il Golfo di Taranto è caratterizzato da una complessa morfologia e da un sistema di canyon sottomarini che costituiscono l'habitat ideale per la ricerca di cibo e per la riproduzione dei cetacei.

La presenza di cetacei e il loro stato di conservazione è un indicatore chiave nella valutazione della qualità dell'ecosistema marino in cui si trovano.





# Introduzione

---

L'obiettivo di questo lavoro di tesi è stato lo sviluppo di modelli predittivi di regressione in grado di stimare l'abbondanza di cetacei nell'area di studio.

Sono state prese in considerazione variabili morfologiche e climatiche per comprendere se e come la formazione geografica del territorio, la pressione antropica e il cambiamento climatico impattano sulla distribuzione dei cetacei.

# Descrizione del data set

Il data set è composta dai dati relativi agli avvistamenti di Stenella Striata (S), Tursiope (T) e Grampo (G), avvenuti dal 2009 al 2019 nel Golfo di Taranto. Le feature morfologiche e climatiche collezionate sono:

- Temperatura
- Densità
- Salinità
- Nitrati
- Fosfati
- Produzione primaria
- Clorofilla – a



JONIAN DOLPHIN CONSERVATION  
ENVIRONMENTAL PROTECTION

# Analisi statistica dei dati

Confrontando i boxplot realizzati per ogni feature rispetto alle tre specie, è emerso un comportamento differente tra G e S – T.

Questo comportamento è stato confermato dal t-test eseguito per ogni possibile coppia di specie.

Feature	G vs S	G vs T	T vs S
N individui	1 / 1	1 / 1	1 / 1
Latitudine e longitudine	2 / 2	2 / 2	2 / 2
Temperatura	13 / 16	14 / 16	6 / 16
Salinità	16 / 16	12 / 16	10 / 16
Densità	14 / 16	16 / 16	8 / 16
Clorofilla-a	1 / 1	1 / 1	1 / 1
Produzione primaria	9 / 16	8 / 16	6 / 16
Nitrati	15 / 16	16 / 16	5 / 16
Fosfati	16 / 16	14 / 16	3 / 16

# Analisi statistica dei dati

Successivamente, abbiamo calcolato il coefficiente di correlazione di Pearson tra il numero di individui e tutte le altre feature, per ogni data set.

STG		T	
Feature	Coeff. Corr.	Feature	Coeff. Corr.
species	-0,2	salinity_400m	-0,283
density_400m	0,137	salinity_300m	-0,283
density_500m	0,133	density_600m	0,226
S		G	
density_400m	0,232	temp_200m	-0,541
density_300m	0,224	density_30m	-0,469
density_200m	0,186	density_20m	-0,464





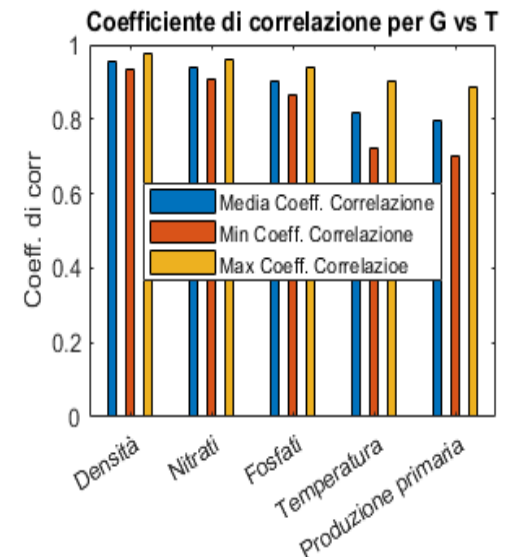
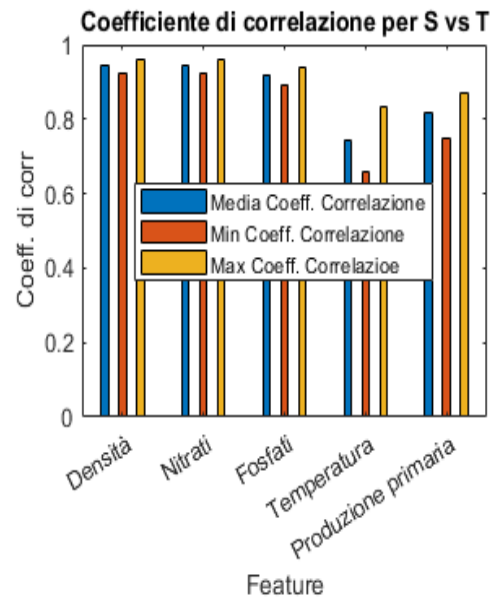
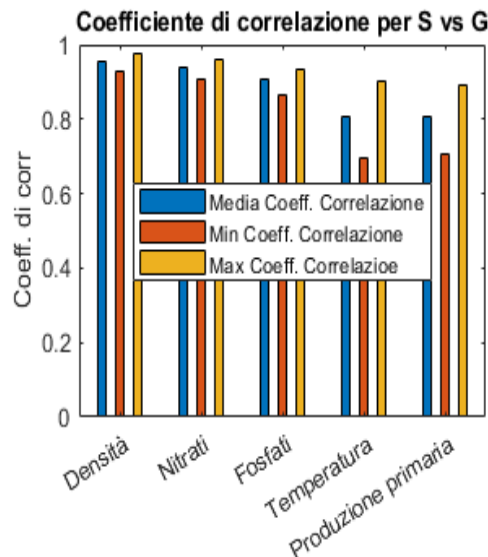
# Analisi statistica dei dati

Dopo aver impacchettato le feature rilevate a differenti profondità, è stato calcolato il coefficiente di correlazione di Pearson con la profondità di rilevamento, per ogni dataset a nostra disposizione.

Feature	STG	T	S	G
Nitrati	0,947	0,946	0,948	0,950
Densità	0,936	0,931	0,937	0,931
Fosfati	0,873	0,868	0,873	0,898
Temperatura	-0,648	-0,644	-0,646	-0,695
Produzione primaria	-0,480	-0,470	-0,481	-0,516
Salinità	0,103	0,114	0,104	0,056

# Analisi statistica dei dati

Per ogni pacchetto di feature è stato calcolato il coefficiente di correlazione di Pearson tra coppie di specie diverse.





# Analisi dei modelli di regressione

---

I modelli di regressione adottati sono stati:

- Random Forest
- LSBoost
- Rete neurale interamente connessa con 2 strati nascosti
- Rete neurale interamente connessa con 3 strati nascosti

Tutti i modelli sono stati addestrati con tuning automatico dei parametri e k-fold cross validation, con  $k = 5$ . I modelli sono valutati in termini di RMSE.

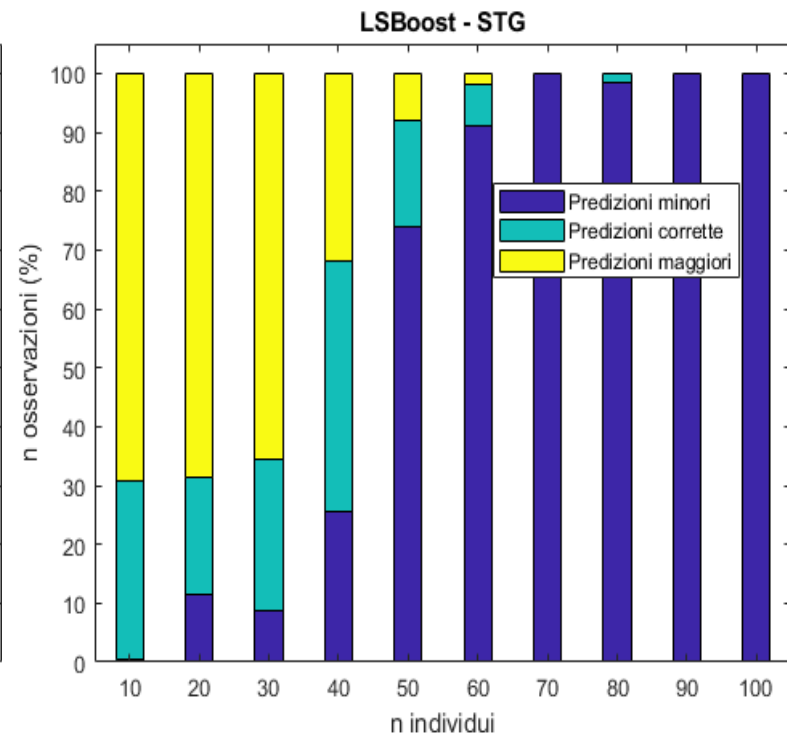
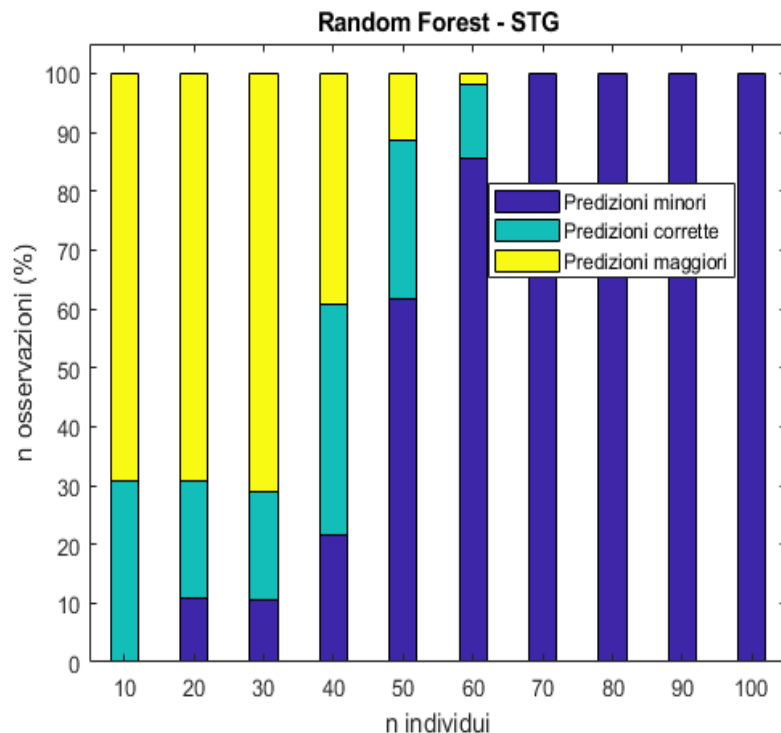
# Discussione dei risultati

I modelli di regressione sono stati addestrati separatamente su tutti e quattro i data set e i risultati sono riportati in tabella.

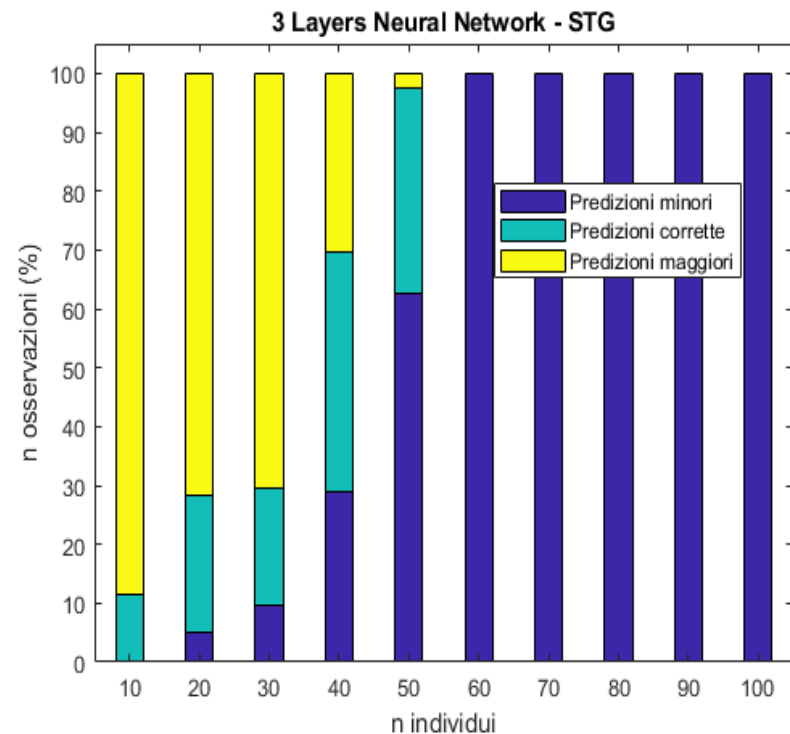
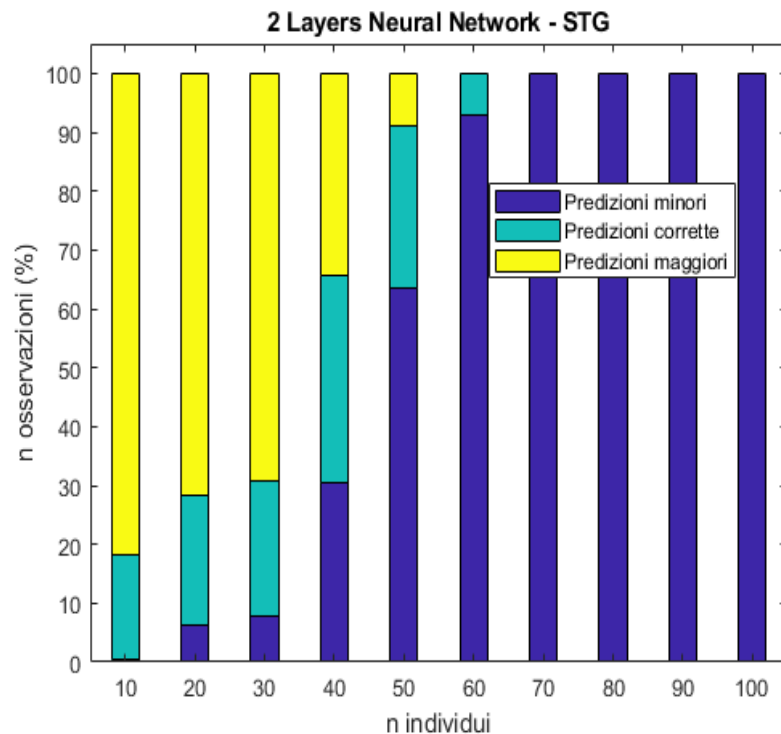
Dataset	Random Forest	LSBoost	Rete neurale a 2 strati	Rete neurale a 3 strati
STG	26,11	25,73	26,12	26,34
S	28,86	28,74	29,34	28,87
T	6,42	6,14	6,46	6,56
G	9,51	7,69	7,61	9,24

# Discussione dei risultati

- Predizioni minori =  $\text{Predetto} < \text{Osservato} - 5$
- Predizioni corrette =  $\text{Predetto} \in [\text{Osservato} - 5; \text{Osservato} + 5]$
- Predizioni maggiori =  $\text{Predetto} > \text{Osservato} + 5$

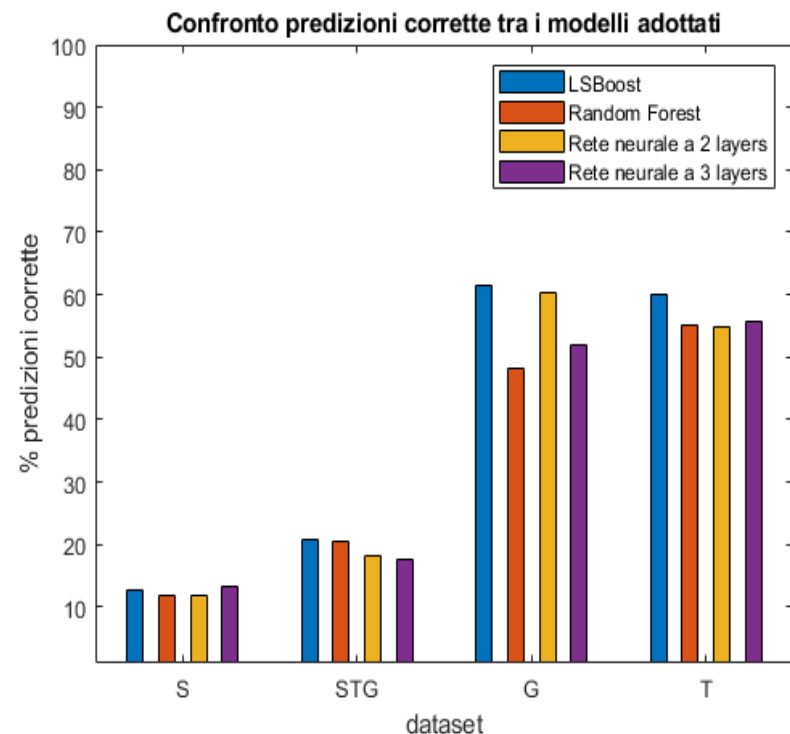
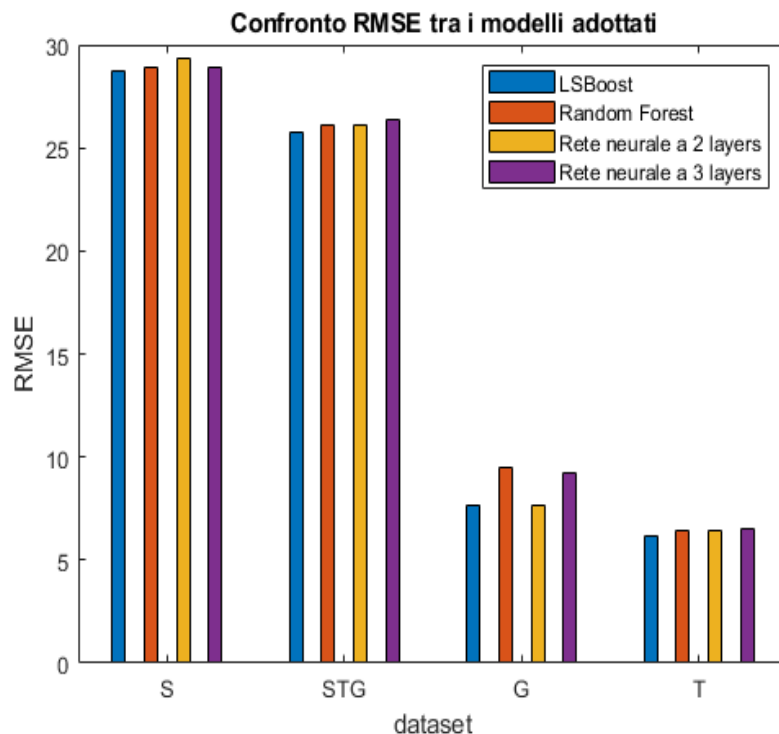


# Discussione dei risultati



# Discussione dei risultati

Infine, sono stati confrontati i risultati ottenuti dai 4 modelli implementati in termini di RMSE e di predizioni corrette.





# Conclusioni e sviluppi futuri

---

Concludendo, lo studio condotto ha dimostrato l'importanza e la significatività delle feature morfologiche e climatiche nella stima di abbondanza dei cetacei.

Inoltre, l'analisi condotta ha mostrato l'esistenza di una correlazione tra le feature a nostra disposizione e il numero di individui osservati.

I risultati ottenuti sono accettabili considerando la natura del problema, confermando inoltre le conoscenze dei biologi.





# Conclusioni e sviluppi futuri

---

Sviluppi futuri del lavoro prevedono

1. l'inserimento nei dataset degli avvistamenti avvenuti dopo il 2019
2. l'acquisizione di nuove feature, migliorando potenzialmente i risultati ottenuti con i modelli sviluppati.



UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO

**STIIMA**   
Sistemi e Tecnologie Industriali Intelligenti  
per il Manifatturiero Avanzato  
Consiglio Nazionale delle Ricerche

# Grazie per l'attenzione