

ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA - CAMPUS DI CESENA

Scuola di Ingegneria e Architettura  
Corso di Laurea Triennale in Ingegneria e Scienze Informatiche

# **IOT PER L'INQUINAMENTO ACUSTICO: algoritmi di machine learning per la calibrazione di microfoni low cost**

**Relatore:**  
**Chiar.mo Prof. Silvia Mirri**

**Presentata da:**  
**Mattia Vincenzi**

**Correlatore:**  
**Dott. Lorenzo Monti**

SESSIONE II  
ANNO ACCADEMICO 2018-2019

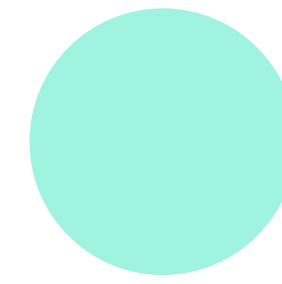
# INQUINAMENTO ACUSTICO

## DEFINIZIONE ED EFFETTI SULLA SALUTE

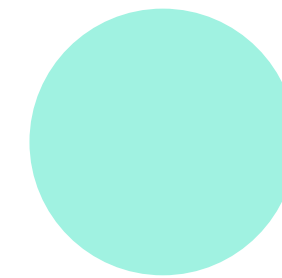
Nella vita quotidiana l'essere umano è sottoposto ad ogni genere di suono desiderato o indesiderato che sia e l'insieme di tutti quei suoni che accompagnano un'attività indesiderata vengono considerati rumore, e di conseguenza inquinamento acustico. Tale piaga non provoca solamente un senso di irritazione e fastidio, ma va ad intaccare anche aspetti sociali e lavorativi, comportando quindi una vasta gamma di disturbi extra-uditivi (ormonali, vascolari, psicologici).



# SCOPI E OBIETTIVI



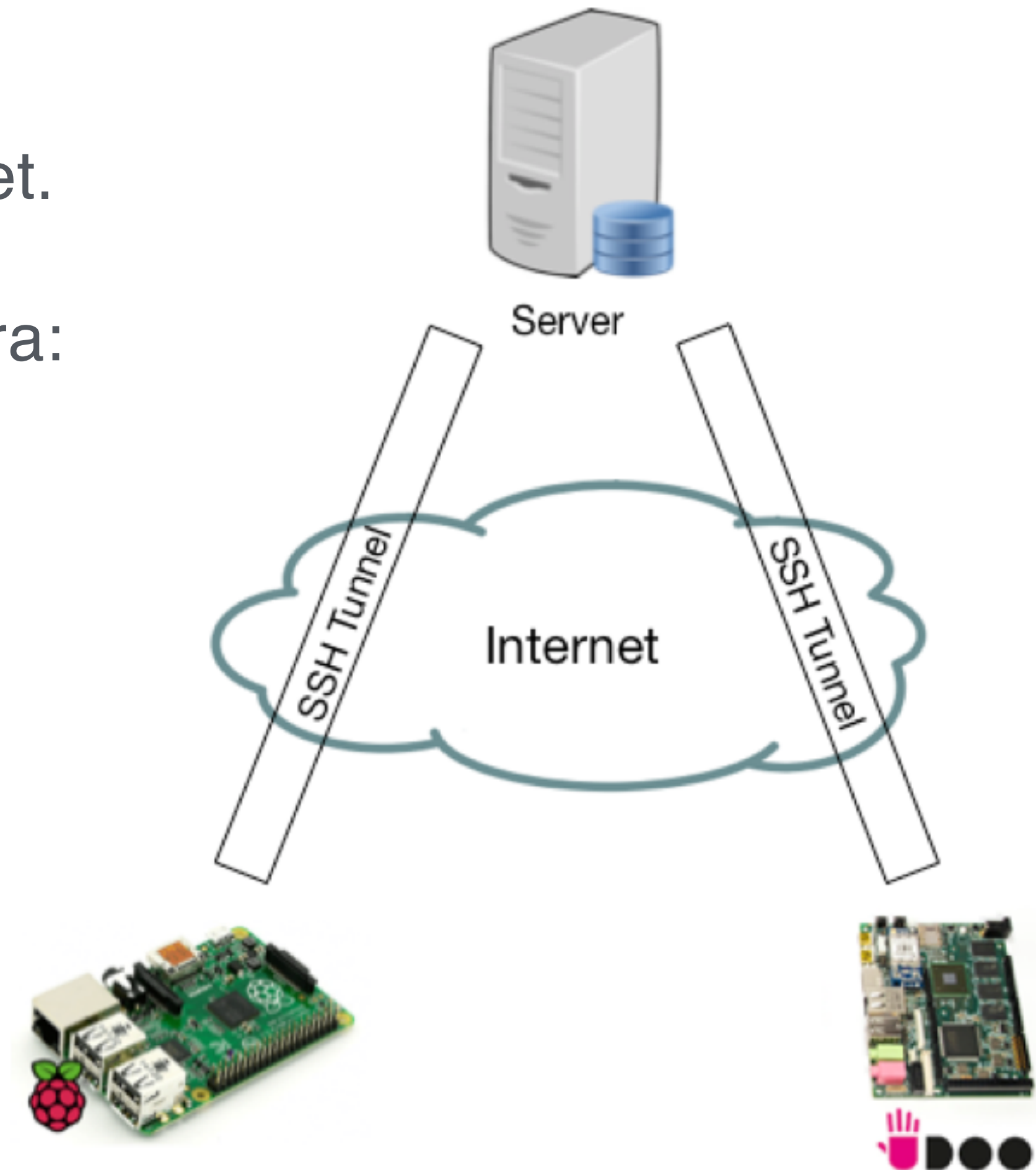
Per poter monitorare il rumore ambientale all'interno dell'**Università degli Studi di Bologna - Campus di Cesena**, al fine di determinare la presenza di inquinamento acustico durante l'orario di studio/lavoro, è necessario avere a disposizione dispositivi che forniscano misure quanto più accurate possibili.



Obiettivo di questa tesi diventa quindi quello di calibrare dei microfoni a basso costo mediante algoritmi di machine learning, cercando di raggiungere le stime ottenute tramite il fonometro.

# ARCHITETTURA PER IL RACCOGLIMENTO DATI

- ▶ Elemento fondamentale per poter applicare strategie di apprendimento automatico è il dataset.
- ▶ Dispositivi Internet Of Things (IoT) nell'architettura:
  - Raspebrry Pi 2 model B + microfono
  - UDOO Neo + fonometro UT351/352
  - Canarin II
- ▶ Questi vengono chiamati edge della rete.



# ARCHITETTURA UTILIZZATA

COMPONENTI CONFIGURATE SUGLI EDGE



Autossh

Crontab

Script per  
invio  
automatico



# METODO DI ANALISI

A scopo di ricerca sono stati utilizzati tre metodi di analisi differenti, creando dataset diversi, e applicando ad ognuno di essi degli algoritmi di machine learning. In questo modo è stato possibile lavorare con un numero di occorrenze e di features diverse per ogni metodo di analisi applicato.

## PRIMO DATASET

- Regressione univariata.
- Variabile indipendente (X): db microfono
- Variabile dipendente (Y): db fonometro

Datetime	$db_{mic}$	$db_{phon}$
2019-05-23 19:00:07	83.8346078673	52.126464811400005
2019-05-23 19:00:08	81.7981022288	50.595703093599994
2019-05-23 19:00:09	82.92256071359999	49.145507781999996

Periodo di campionamento: un mese (circa un milione di occorrenze)

# METODO DI ANALISI

A scopo di ricerca sono stati utilizzati tre metodi di analisi differenti, creando dataset diversi, e applicando ad ognuno di essi degli algoritmi di machine learning. In questo modo è stato possibile lavorare con un numero di occorrenze e di features diverse per ogni metodo di analisi applicato.

## SECONDO DATASET

- Regressione multivariata
- Frequenza campionamenti al secondo
- Variabile indipendente (X): db microfono, dati ambientali
- Variabile dipendente (Y): db fonometro

DateTime	$db_{mic}$	$db_{phon}$	PM1.0	PM10	PM2.5	Pres	Temp	Hum
19:00:13	83.9	51.0	3.0	7.0	5.0	1007.7	21.9	59.2
22:06:54	81.7	50.7	4.0	6.9	6.9	1008.4	21.4	60.5
23:55:34	96.2	64.8	4.4	7.5	7.5	1008.5	21.1	60.2

Periodo di campionamento: un mese (circa un milione di occorrenze)

# METODO DI ANALISI

A scopo di ricerca sono stati utilizzati tre metodi di analisi differenti, creando dataset diversi, e applicando ad ognuno di essi degli algoritmi di machine learning. In questo modo è stato possibile lavorare con un numero di occorrenze e di features diverse per ogni metodo di analisi applicato.

## TERZO DATASET

- ▶ Regressione multivariata
- ▶ Frequenza campionamenti al minuto
- ▶ Variabile indipendente (X): db microfono, dati ambientali
- ▶ Variabile dipendente (Y): db fonometro

DateTime	$db_{mic}$	$db_{phon}$	PM1.0	PM10	PM2.5	Pres	Temp	Hum
19:00:13	83.9	51.0	3.0	7.0	5.0	1007.7	21.9	59.2
19:01:14	82.4	50.6	3.0	6.0	5.0	1007.7	21.9	59.2

Periodo di campionamento: due mesi (circa quaranta mila occorrenze)



# RISULTATI MIGLIORI

Ad ognuno dei tre dataset sono stati applicati gli stessi algoritmi di machine learning. Nella tabella sono riportati i risultati riassuntivi del migliore modello per ciascun dataset.

	Miglior Modello	Coef. $R^2$	RMSE	Err. Relativo
Metodo I	Reg. polinomiale	0.85	4.15	5.78%
Metodo II	Random forest	0.93	2.74	2.57%
Metodo III	Random forest	0.87	3.59	3.90%

## VALORI IPERPARAMETRI

1

### METODO I

- Grado polinomio: 20
- Regularizzazione: 0.1

2

### METODO II

- Numero di alberi decisionali: 40

3

### METODO III

- Numero di alberi decisionali: 99

# INTEGRAZIONE MODELLI NEL SISTEMA

INTEGRAZIONE DEI MODELLI NEL TOOL DI MONITORAGGIO  
DEL RUMORE INSTALLATO SU DISPOSITIVI EMBEDDED:

## Calibrazione Tiny

- Leggera
- Semplice da applicare
- Sfrutta minor numero di componenti possibile
- Si basa sul modello ottenuto dal Metodo I

## Calibrazione Full

- Precisa
- Complessa da applicare
- Utilizza anche i dati raccolti dal Canarin II
- Si basa sul modello ottenuto dal Metodo II

# 01

Il modello migliore è quello basato su random forest, ottenuto dal secondo dataset ed addestrato utilizzando 40 alberi decisionali.

# 02

Nel sistema è stata integrata la **calibrazione tiny** in quanto non necessita dell'aggiunta di componenti.

# 03

Il modello pre-addestrato sul primo dataset è stato esportato in binario, venendo quindi richiamato ad ogni campionamento del microfono.

## CONCLUSIONE

# SVILUPPI FUTURI

- Aumento valore iperparametri
- Utilizzo di algoritmi di learning più complessi
- Campionamento delle features ambientali con frequenza al secondo







SESSIONE INVERNALE  
ANNO ACCADEMICO  
2018-2019

# THE END

Presentata da; MATTIA VINCENZI