

POLITECNICO DI TORINO  
Tesi di Laurea Magistrale

---

# Station-level forecasting in bike sharing systems using data mining techniques

---

**Relatori:**

*Prof.ssa* Silvia Anna Chiusano

*Prof.ssa* Tania Cerquitelli

**Candidato:**

Khalid Benihich

# Sommario

---

- Bike Sharing Systems
- Predizione mediante Data Mining
- Formalizzazione della procedura
- Validazione sperimentale
- Implementazione del framework
- Applicazione mobile

# Bike Sharing Systems

---

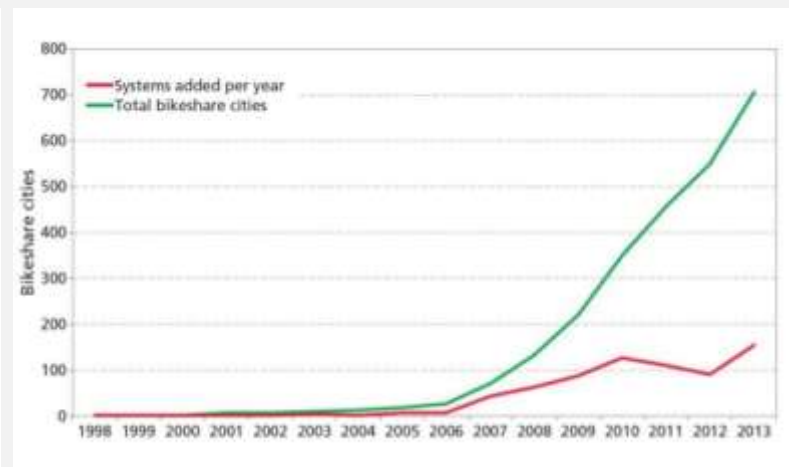
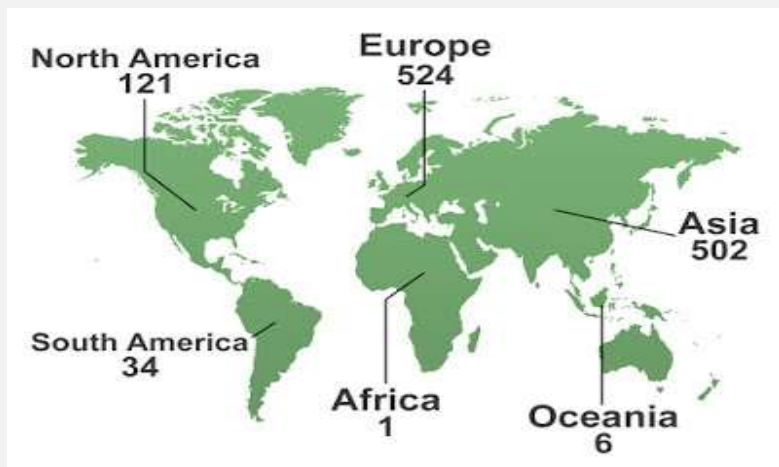


## Vantaggi:

- Riduzione traffico auto e riduzione inquinanti
- Maggior connessione tra altri mezzi pubblici
- Effetti benefici salute degli utilizzatori
- Riduzione superfici dedicate al parcheggio
- Basso costo

# Diffusione e crescita

---



# Funzionamento subottimale

---



# Funzionamento subottimale

---



€€€

# Data Mining

---

# Classificazione

---

- Dati di Training: set di record  $(x, y)$ 
  - $x$  = **attribute set** (valori discreti e continui)
  - $y$  = **etichetta di classe** (solo valori discreti)

La classificazione è l'operazione apprendimento della **funzione target** che mappa **x** con l'etichetta di classe **y**

*funzione target = modello di classificazione*



# Classificatori Naive Bayes

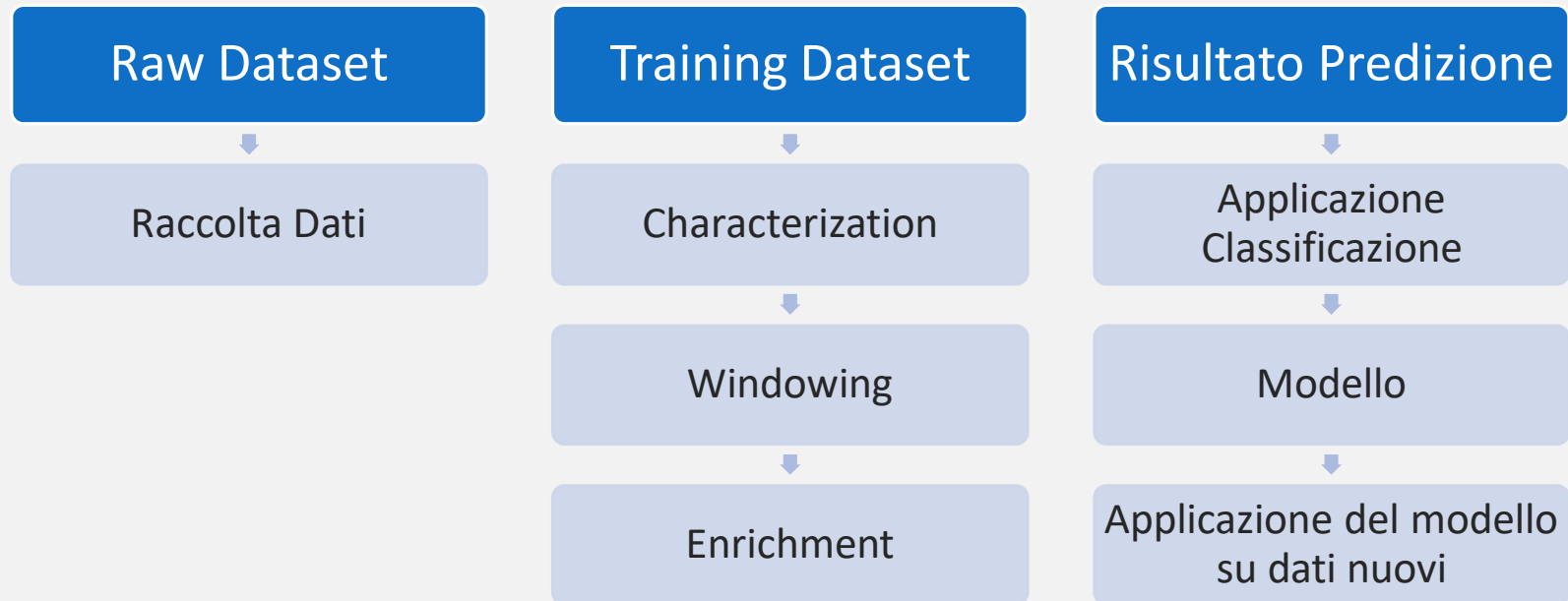
---

# Principali Algoritmi

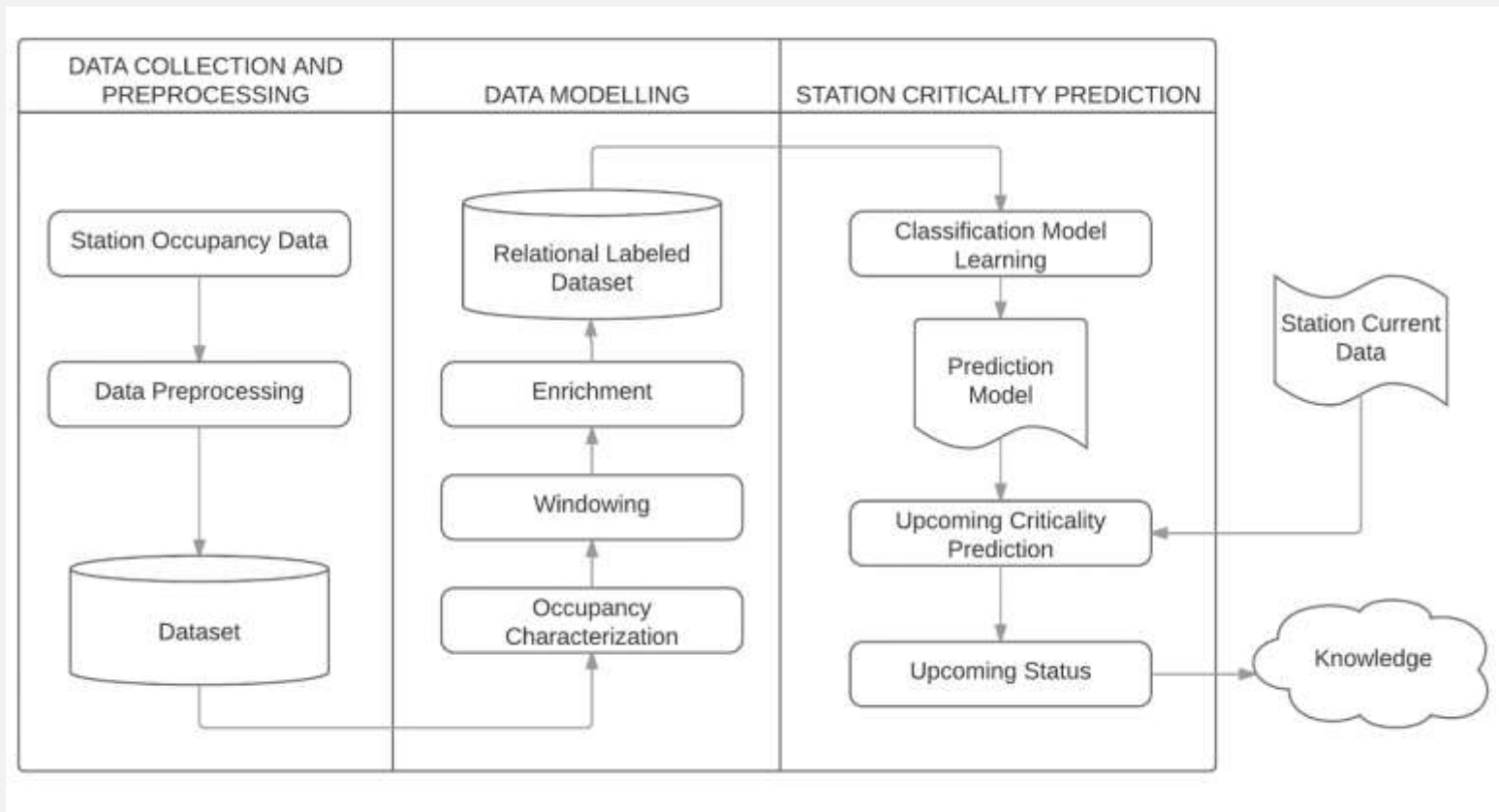
---

# Processo

---



# Processo



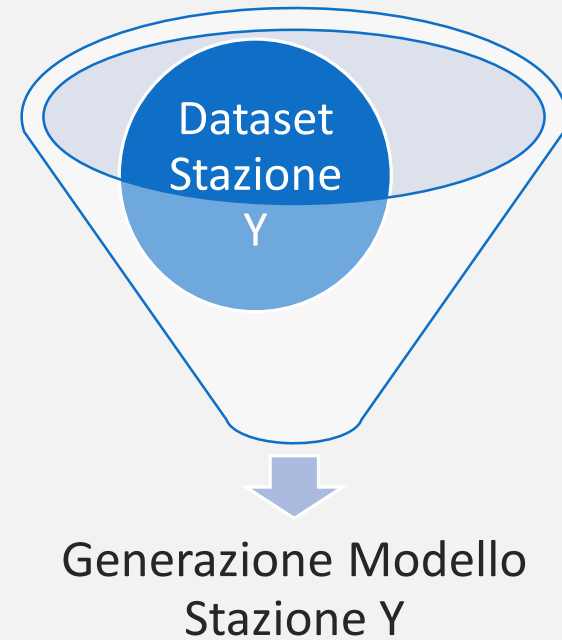
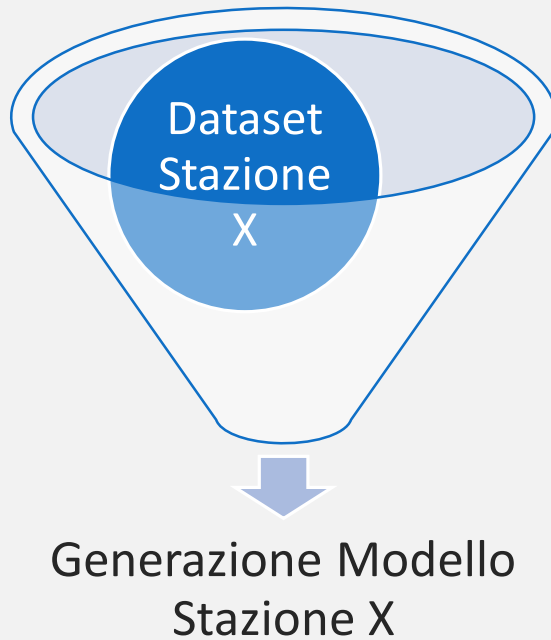
# Prove Sperimentali

---

- Applicazione del processo su dati reali della rete di **New-York**:
  - **8,4** milioni di abitanti
  - **10000** biciclette
  - **600** stazioni
- Tre differenti periodi dell'anno
  - Aprile - Maggio
  - Giugno - Luglio
  - Settembre - Ottobre
- Tramite tre approcci differenti nella creazione dei modelli

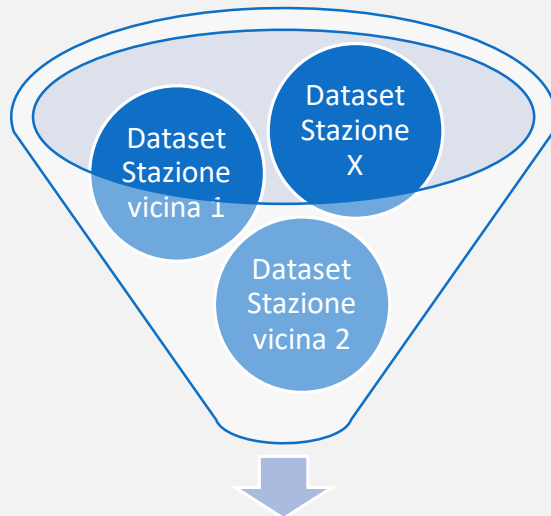
# Singola stazione

---

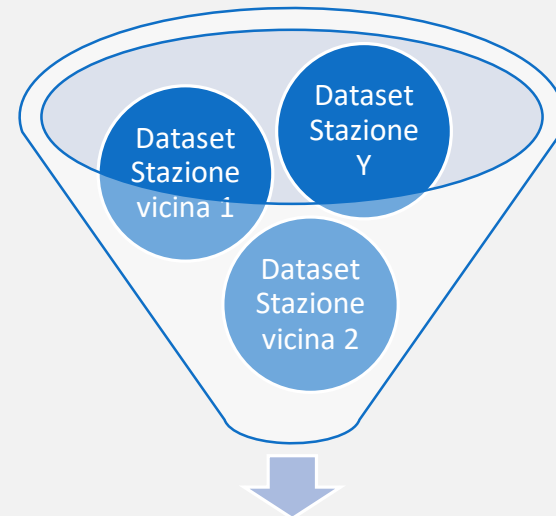


# Singola Stazione più i Vicini

---



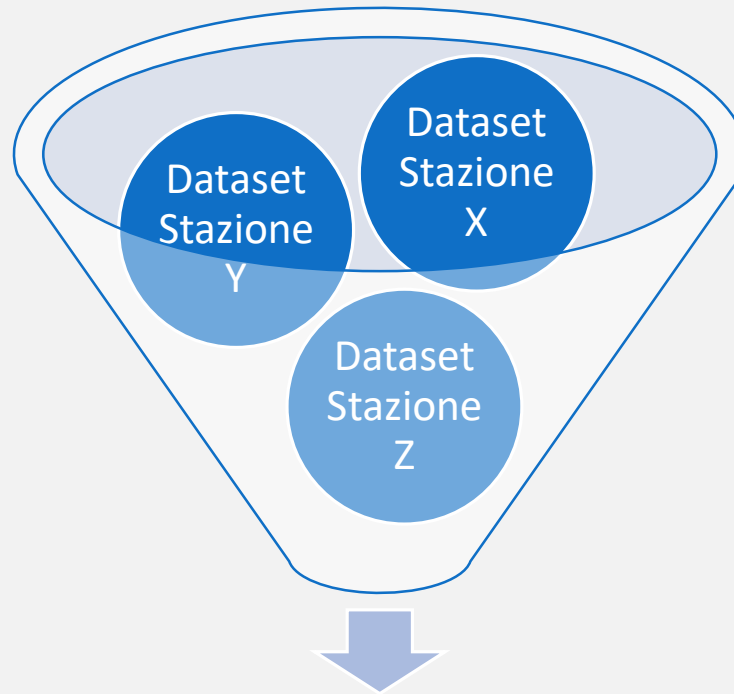
Generazione Modello  
Stazione X



Generazione Modello  
Stazione Y

# Intera Rete

---



Generazione Modello Unico



# Valutazione dei Modelli

---

		Predicted	
		Class = 1	Class = 0
Actual	Class = 1	$f_{11}$ (TP)	$f_{10}$ (FN)
	Class = 0	$f_{01}$ (FP)	$f_{00}$ (TN)

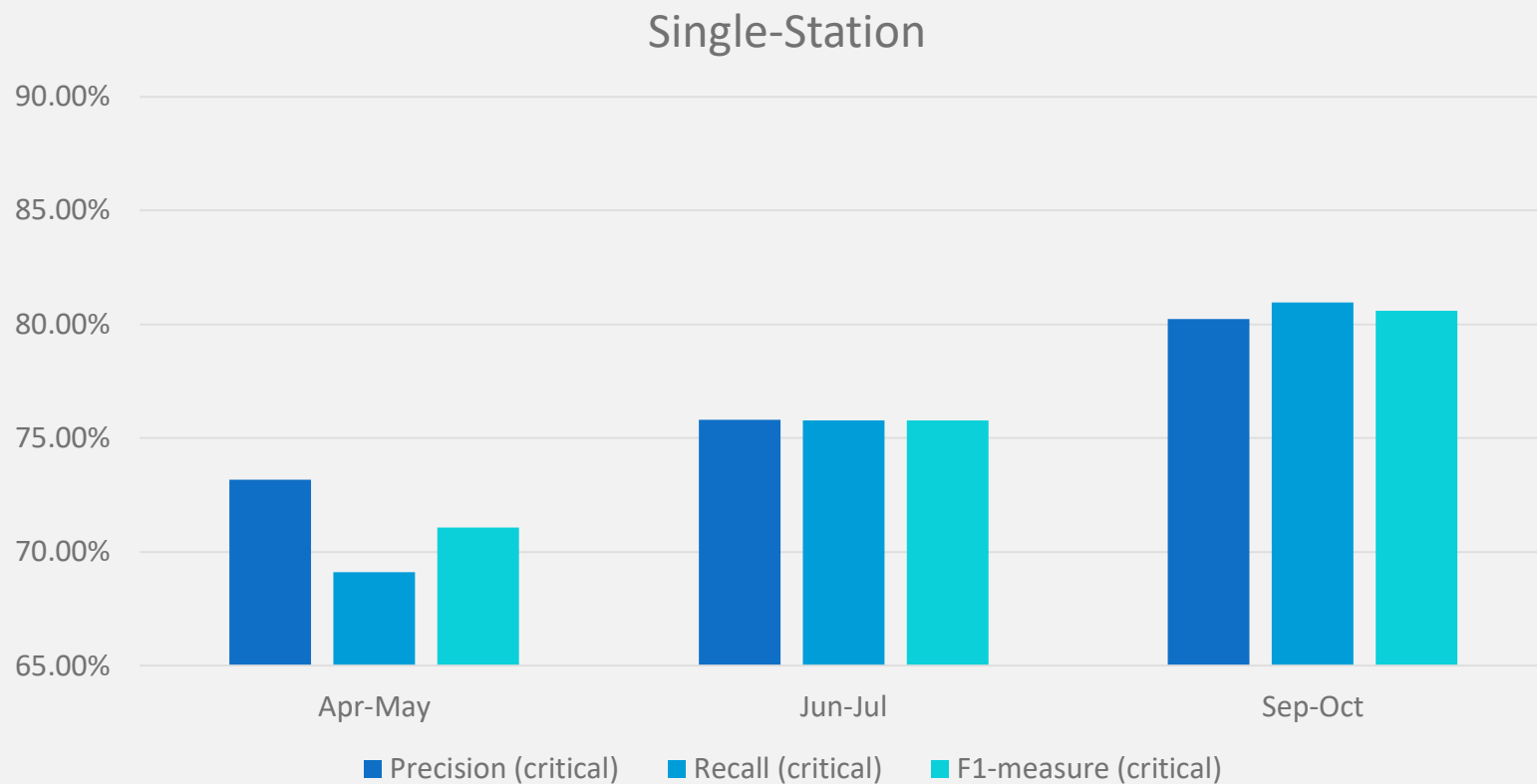
$$Precision, a = \frac{TP}{TP + FP}$$

$$Recall, r = \frac{TP}{TP + FN}$$

$$F1 - measure = \frac{1}{\frac{1}{p} + \frac{1}{r}}$$

# Risultati 1° Approccio

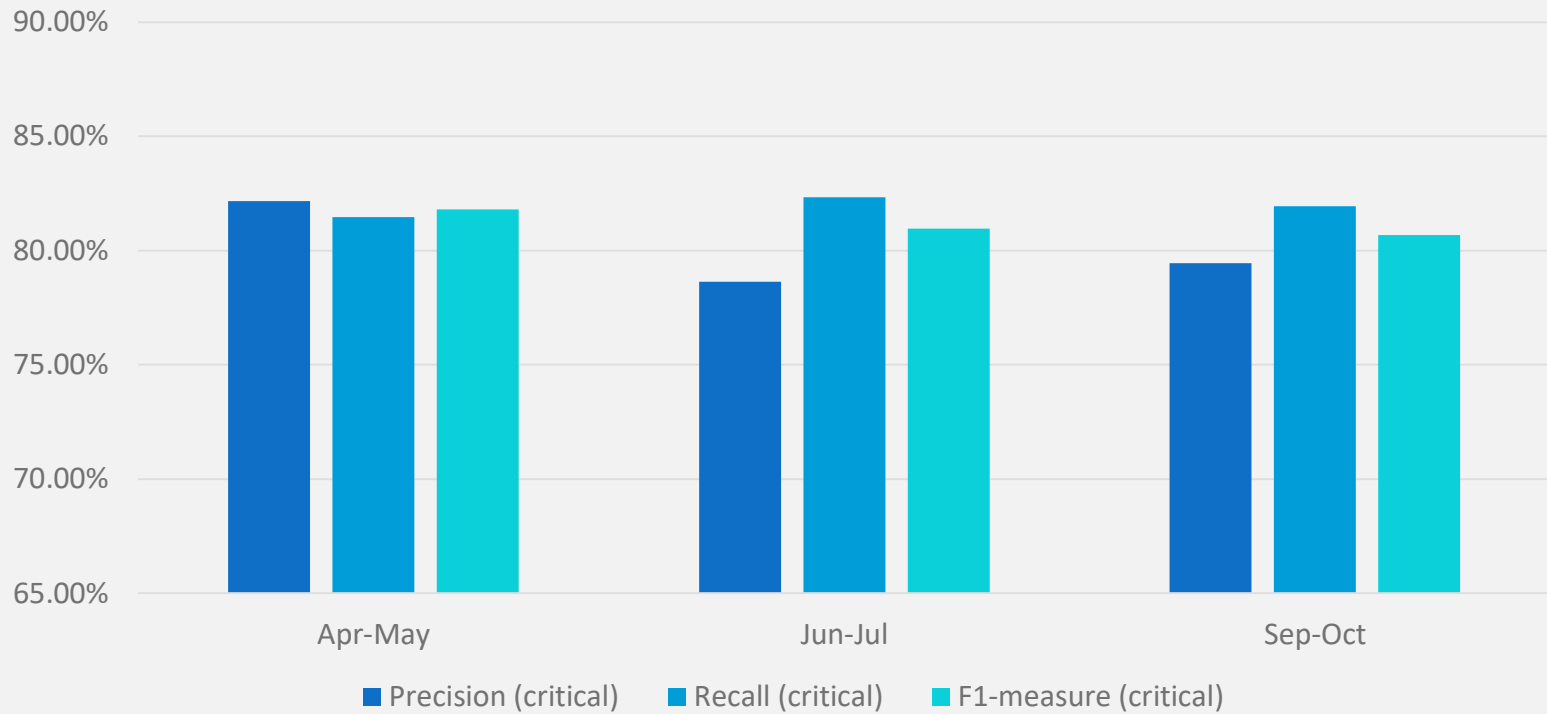
---



# Risultati 2° Approccio

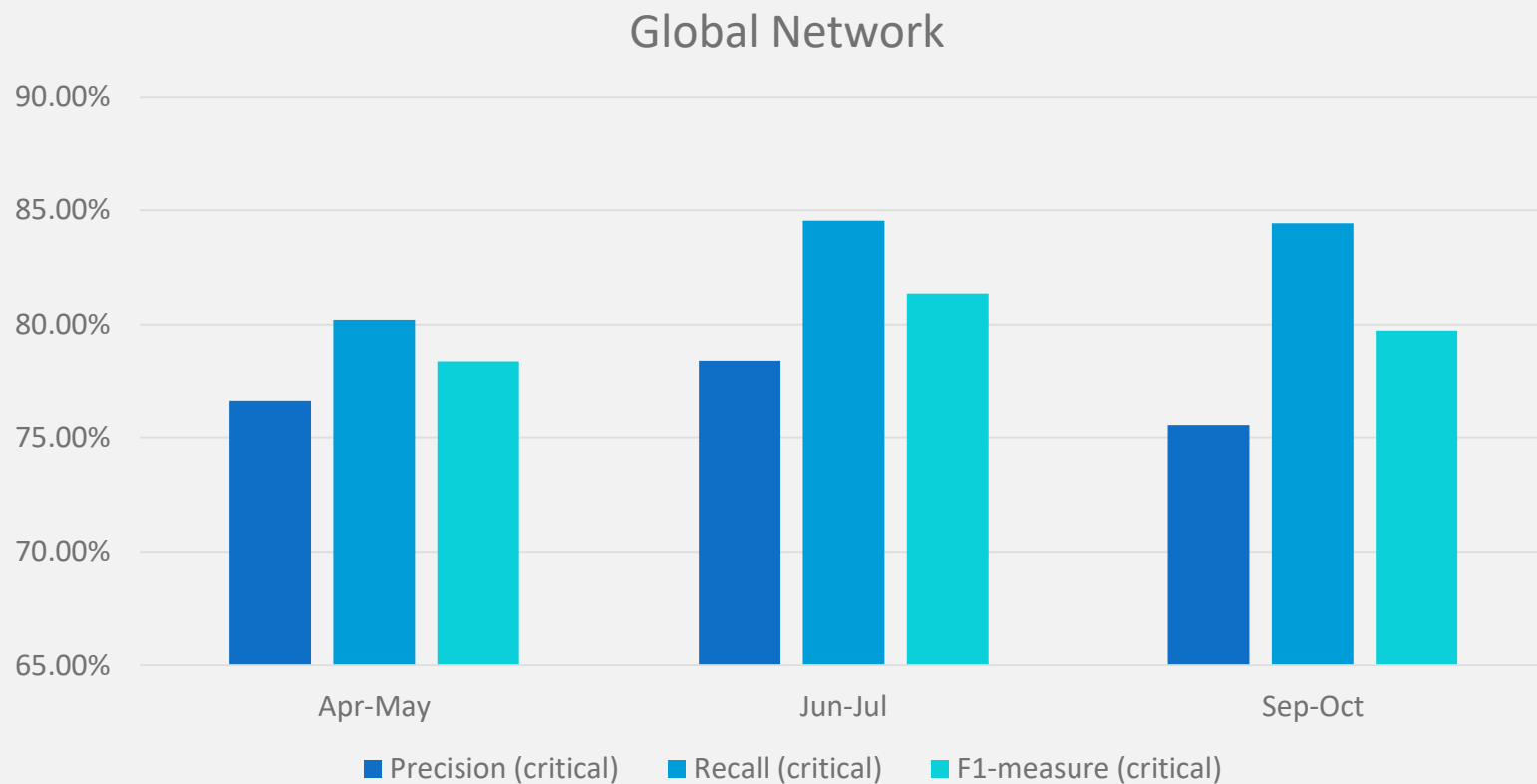
---

## Single-Station & Neighbors (2 km)



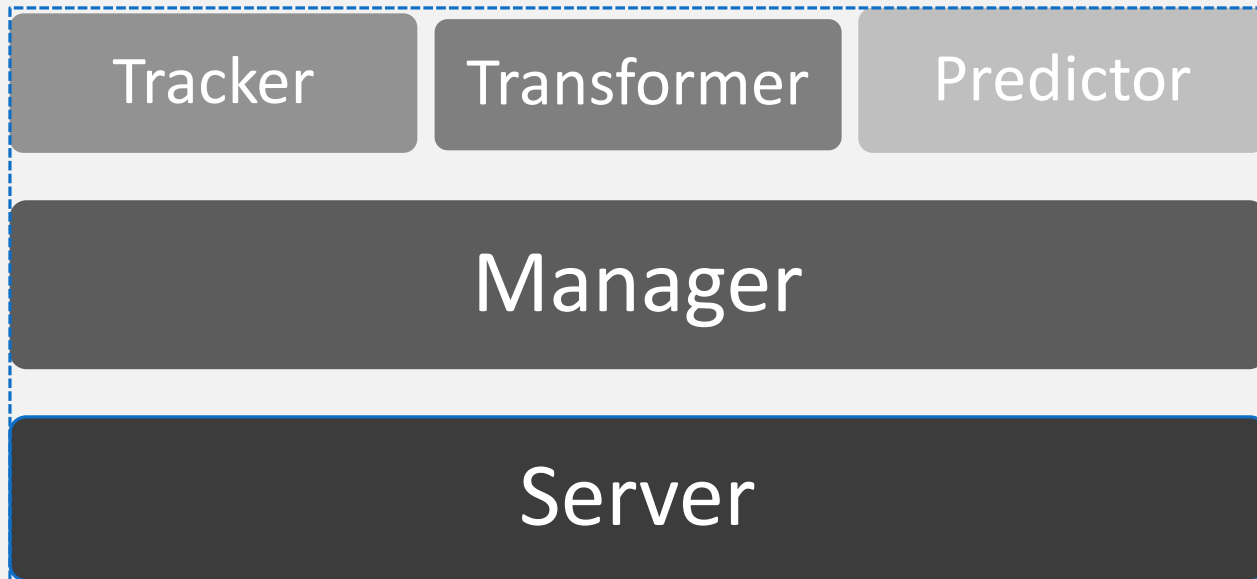
# Risultati 3° Approccio

---



# Implementazione Framework

---



# Caratteristiche tecniche

---

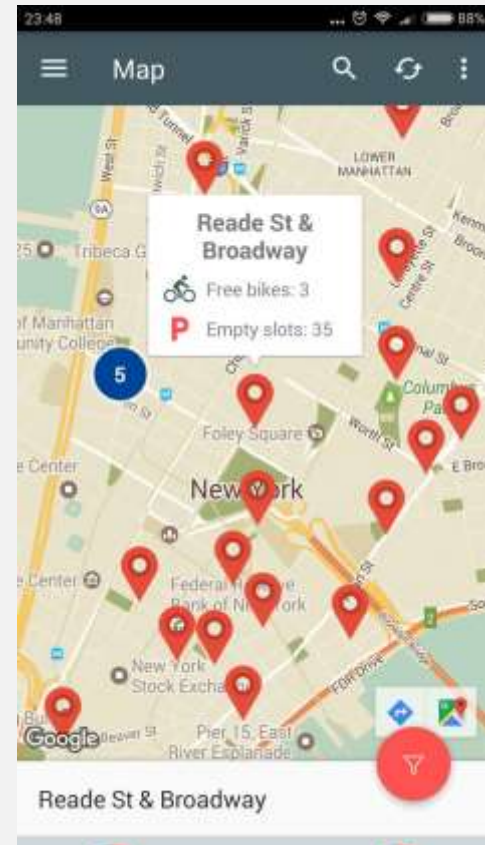
- Linguaggio di programmazione Java
  - Portabilità
- API RESTFUL per le comunicazioni client server
  - Adatto alla struttura modulare
- Contenitore Apache
  - Permette di far girare il framework come web app

# Applicazione Mobile

---

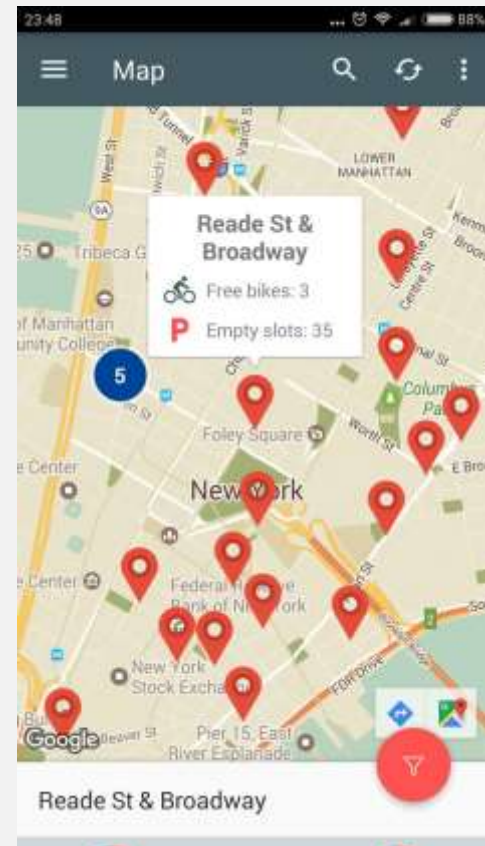
- Scelta piattaforma mobile
  - Android
- Information Visualization
- Struttura semplice
  - Solo due Activity

# Funzionamento App



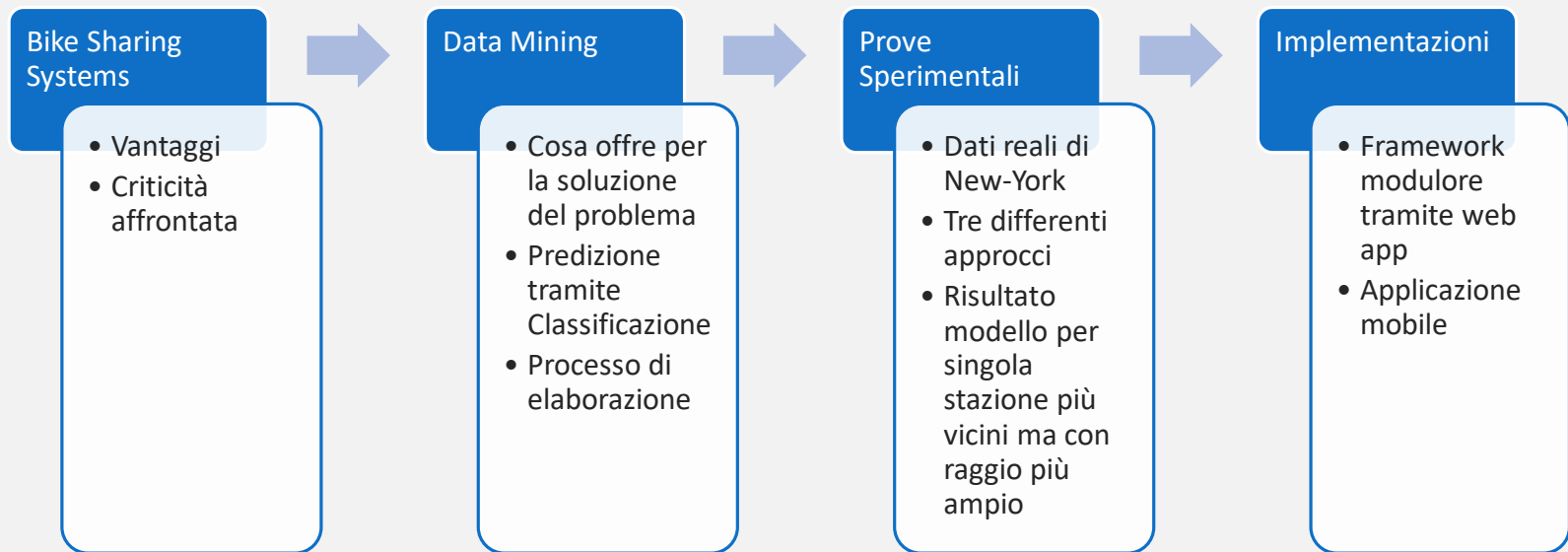


# Funzionamento App



# Riepilogo

---



# Sviluppi Futuri

---

- Verifica dell'approccio «Single-Station & Neighbors» su **altre città**
- Integrazione di dati **meteo** e del **traffico** nella creazione dei modelli
- Estendere il **framework** per soluzione di **problemi simili**
  - Es. Consumi energetici, Car sharing, ecc.

---

Grazie per l'attenzione

---