

# DATA MANAGEMENT DATA VISUALIZATION PROJECT

MATTIA VENTOLA  
ALESSANDRO MOTTA

**Social listening della finale di  
Champions League:  
PSG – Bayern Monaco**

## **ANALISI DEL PROBLEMA**

Introduzione al tema e presentazione  
del problema di business al quale  
si è voluto rispondere

# 01

## **DATA ACQUISITION**

Kafka e Web Scraping

# 02

## **DATA STORAGE**

Mongo DB & NiFi

# 03

# CONTENUTI

# 04

## **PRE-PROCESSING & DATA QUALITY**

# 05

## **CONCLUSIONI e CRITICITÀ**



# 01

## ANALISI DEL PROBLEMA

---



# DATI:

**TEMA:** SENTIMENT ANALYSIS DEI TWEET RIGUARDANTI LA FINALE DI CHAMPIONS:  
PSG VS BAYERN MONACO.  
CONFRONTO CON LE RELATIVE PRESTAZIONI IN CAMPO DEI GIOCATORI.

# OBIETTIVO:

ANALISI DEL TRAFFICO TWITTER DELLA PARTITA IN RELAZIONE AGLI EVENTI  
ACCADUTI. ANALISI DELLA RELAZIONE TRA LA SENTIMENT ANALYSIS DEI  
TWEET E LE PRESTAZIONI SPORTIVE DEI CALCIATORI.



02

# DATA ACQUISITION

# KAFKA

PRODUCER 1

CONSUMER 1

PRODUCER 2

CONSUMER 2



# WEB SCRAPING

WebScraper.io

The screenshot displays the WebScraper.io web scraper interface. The main content area shows a table of Paris Saint-Germain players and their statistics. The table has columns for player name, position, and various statistics. The interface includes a sidebar with navigation options like 'Elements', 'Console', 'Sources', 'Network', 'Performance', 'Memory', 'Application', 'Security', 'Lighthouse', and 'Web Scraper'. The 'Web Scraper' tab is active, showing a configuration panel with fields for 'Id', 'Type', 'Selector', 'Multiple', 'Regex', and 'Parent Selectors'. The 'Id' field is set to 'stats', 'Type' is 'Text', 'Selector' is 'p', 'Multiple' is checked, 'Regex' is 'mpix', and 'Parent Selectors' is 'root'.

	Goalkeepers	Defence	Midfielders	Forwards	Goalkeepers	Defence	Midfielders	Forwards	Goalkeepers	Defence	Midfielders	Forwards
1	Kelvin Nolasco	33, FCB	0	0	0	0	0	0	0	0	0	0
2	Thilo Kehrer	23, DFB	4	0	0	1	1	0	0	0	0	0
3	Thiago Silva	35, DFB	1	1	6	1	1	0	0	0	0	0
4	Presnel Kimpembe	25, DFB	0	0	4	1	0	0	0	0	0	0
5	Juan Bernat	27, DFB	0	3	3	0	0	0	0	0	0	0
6	Leandro Paredes	28, DFB	0	2	1	0	1	0	0	0	0	0
7	Marquinhos	26, DFB	3	2	3	1	2	0	0	0	0	0
8	André Herrera	21, DFB	1	2	0	1	1	0	0	0	0	0

Whoscored.com



03

# DATA STORAGE

---





mongoDB®





# 04

## PRE-PROCESSING & DATA QUALITY

# PRE-PROCESSING

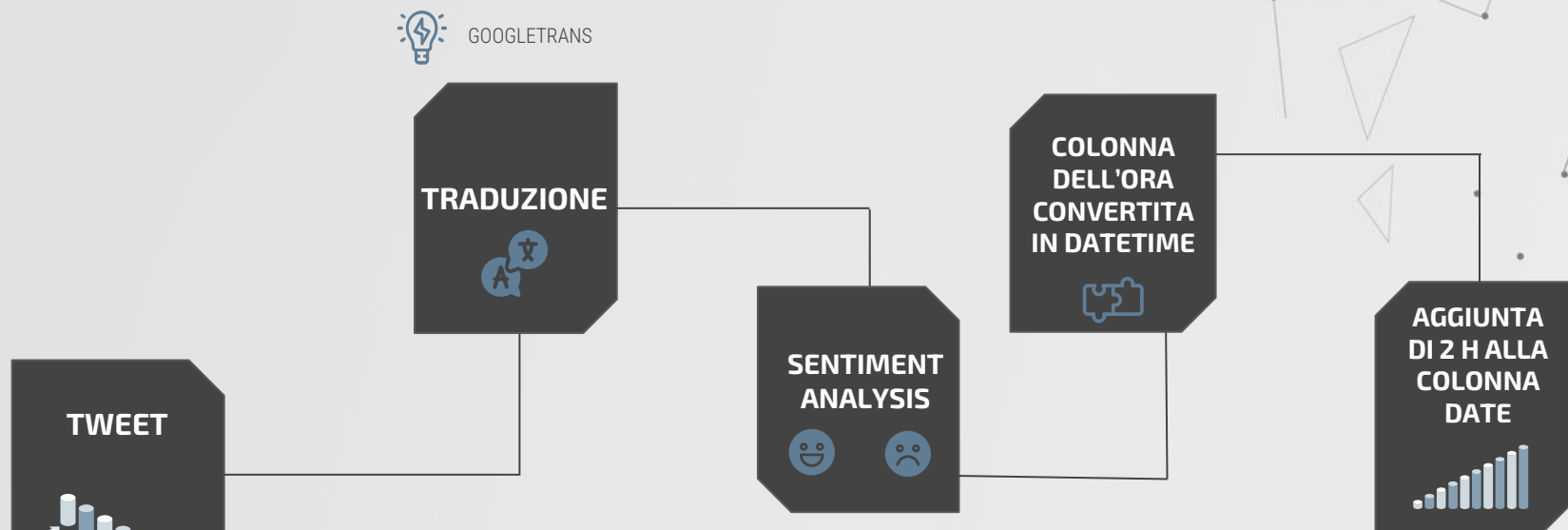


**RICERCA BOT**

**ELIMINAZIONE:**

- EMOJI
- LINK
- @ e #
- PUNTEGGIATURA
- SIMBOLI

# PRE-PROCESSING



# DIVISIONE DATASET

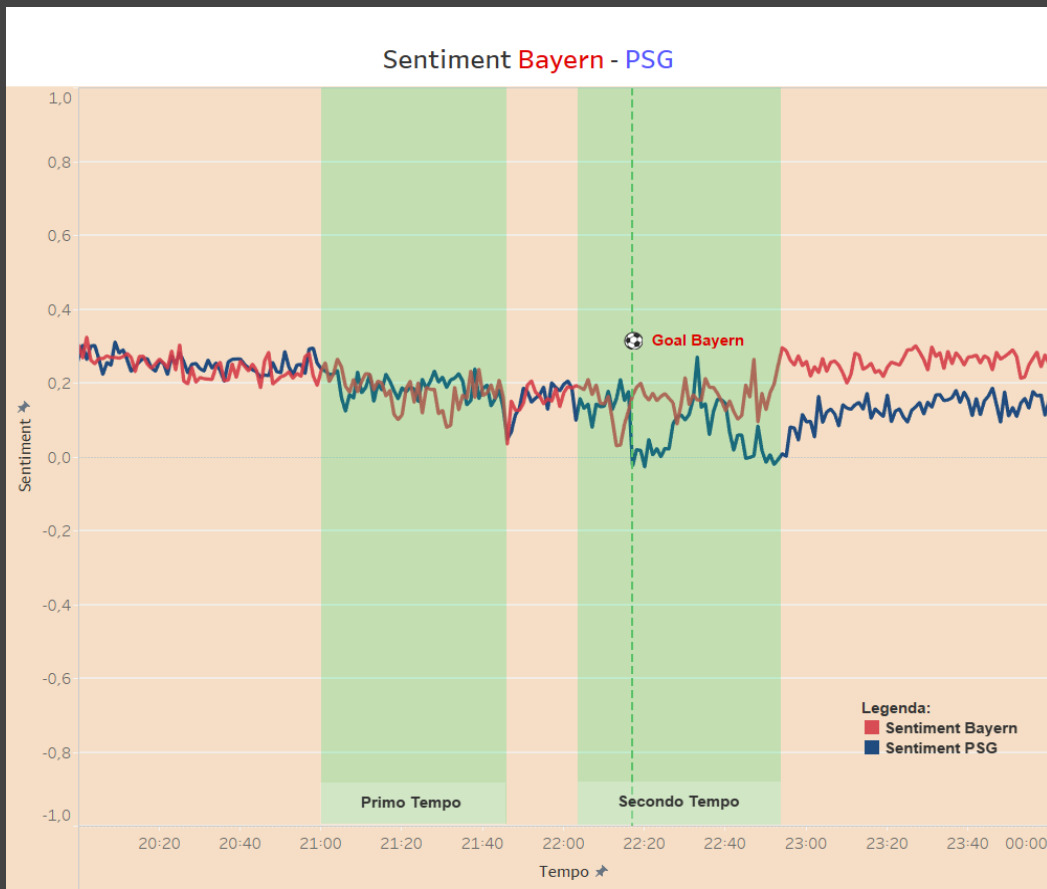
**DATASET SQUADRE  
SELEZIONANDO "PSG"  
E NON "BAYERN" E  
VICEVERSA**

**-DATASET GIOCATORI SELEZIONANDO  
"NOME\_GIOCATORE" ED ESCLUDENDO  
LA LISTA DEGLI ALTRI GIOCATORI**

**- AGGIUNTA NUMERO TWEET, POLARITÀ  
MEDIA E DEVIATIONE STANDARD**

**-CONVERSIONE POLARITÀ [0,10]**

# DIVISIONE DATASET





# CRITICITÀ e CONCLUSIONI

## CRITICITÀ:

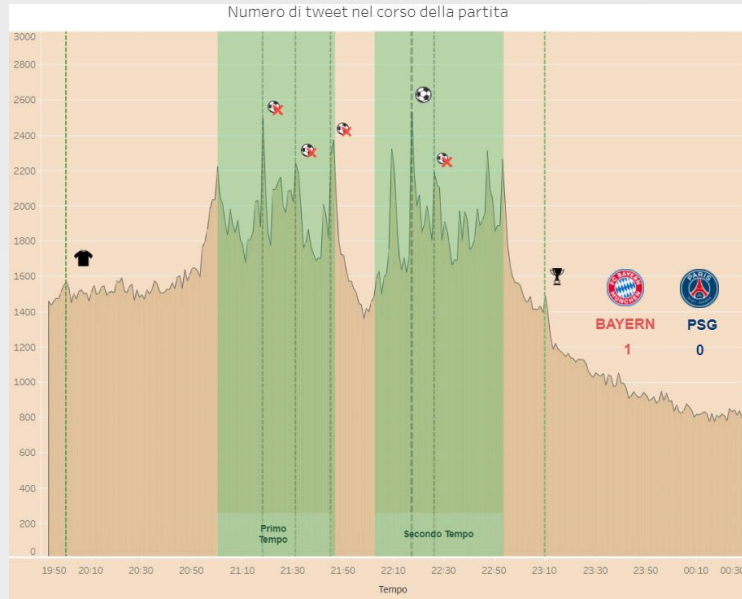
- SVILUPPO CODICE PER LO STREAMING
- CONNESSIONE MACCHINA VIRTUALE
- MIGLIORE IDENTIFICAZIONE BOT

## CONCLUSIONI:

- TRAFFICO TWITTER E SENTIMENT IN LINEA CON EVENTI DEL MATCH
- COERENZA TRA VALORE SENTIMENT E VOTO DEI GIOCATORI

# TABLEAU

<https://public.tableau.com/profile/mattia.ventola>



## Sentiment vs Voto