

Analisi delle catastrofi Sandy e Joplin attraverso il framework SPARK

- Mattia Gatto,
Matricola: 216649
- Francesco Maria Granata,
Matricola: 216648

- Professore:
Paolo Trunfio
- Esercitatore:
Fabrizio Marozzo

INTRODUZIONE

Analisi dei contenuti dei Tweet generati durante due differenti catastrofi

- Sandy
- Joplin

Tweet suddivisi in tre classi principali

- Personale
- Informativo:
 - diretto
 - indiretto
- Altro

01

Utilizzo del
modulo python
pandas

02

Visualizzazione
dei dati in formato
tabellare

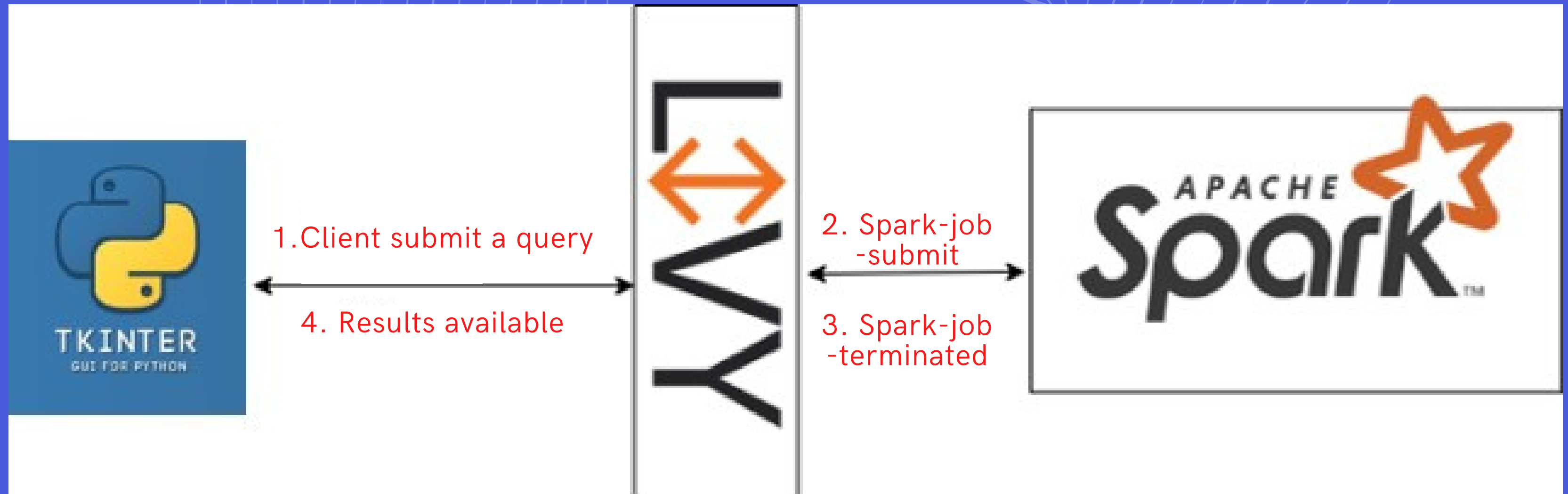
03

Eliminazione delle
colonne a
crowdsourcing

03

Preelaborazione dei dati

FRAMEWORK E SERVIZI UTILIZZATI



Framework Spark



Utilizzato all'interno del progetto per l'analisi dei dataset.

Il linguaggio utilizzato per sfruttare il framework è Scala

Gestendo le dipendenze e la fase di compilazione con il build tool "sbt".

Tipo di analisi realizzata

01

Query che restituiscono statistiche di tipo count sulle classi di Tweet presenti nel Dataset considerato.

02

Query che restituiscono i Tweet di una certa classe oppure gli Utenti autori di tali Tweet

Servizio Livy



Crea un'API REST attraverso la quale il client potrà sfruttare il cluster SPARK.

E' in grado di automatizzare l'esecuzione di job SPARK

Monitora lo stato dei job SPARK che manda in esecuzione.

Metodi per la comunicazione tra client e Livy server

01

“livy_submit_job”
effettua richieste di
tipo POST

02

“livy_job_success” eff
ettua richieste di tipo
GET

01

Utilizzo del modulo
Python TKinter

02

Utilizzo del modulo
Python Pandas

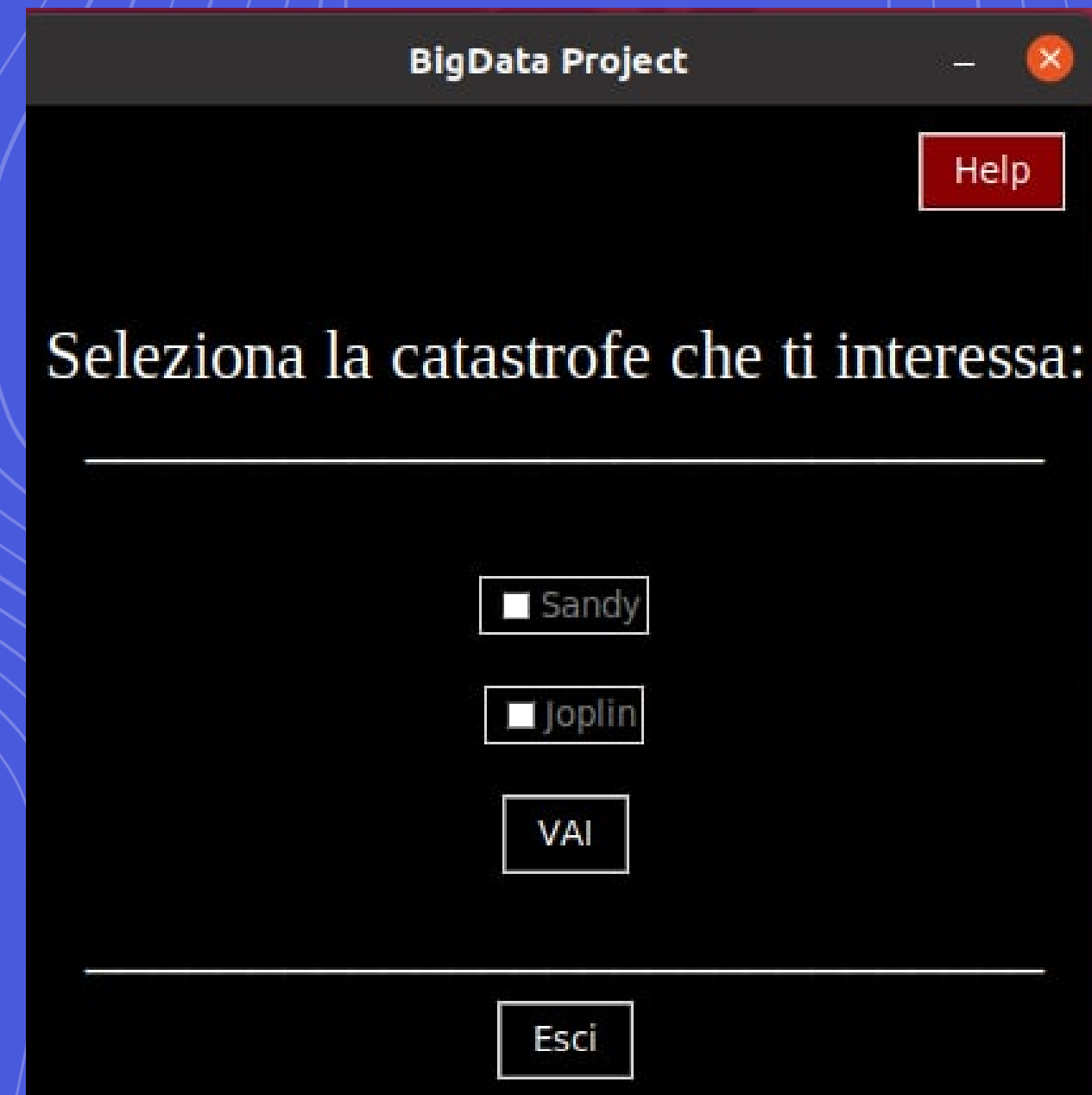
03

Utilizzo del modulo
Python Matplotlib

09

Interfaccia Grafica

ESEMPIO DI ESECUZIONE (1).



A screenshot of a Java Swing window titled "BigData Project". The window has a dark gray title bar with standard window controls (minimize, maximize, close). The main content area is black. In the top right corner, there is a red rectangular button labeled "Help". Below the title bar, the text "Seleziona la catastrofe che ti interessa:" is displayed in a light gray font. Underneath this text is a horizontal white line. Below the line, there are three vertically stacked buttons: "Sandy", "Joplin", and "VAI". At the bottom of the window, there is another horizontal white line, and below it is a button labeled "Esci".

BigData Project

Help

Seleziona la catastrofe che ti interessa:

Sandy

Joplin

VAI

Esci

ESEMPIO DI ESECUZIONE (2)

11

Sandy e Joplin Queries

Confronto tra Sandy&Joplin
Decidi su quale argomento fare query:

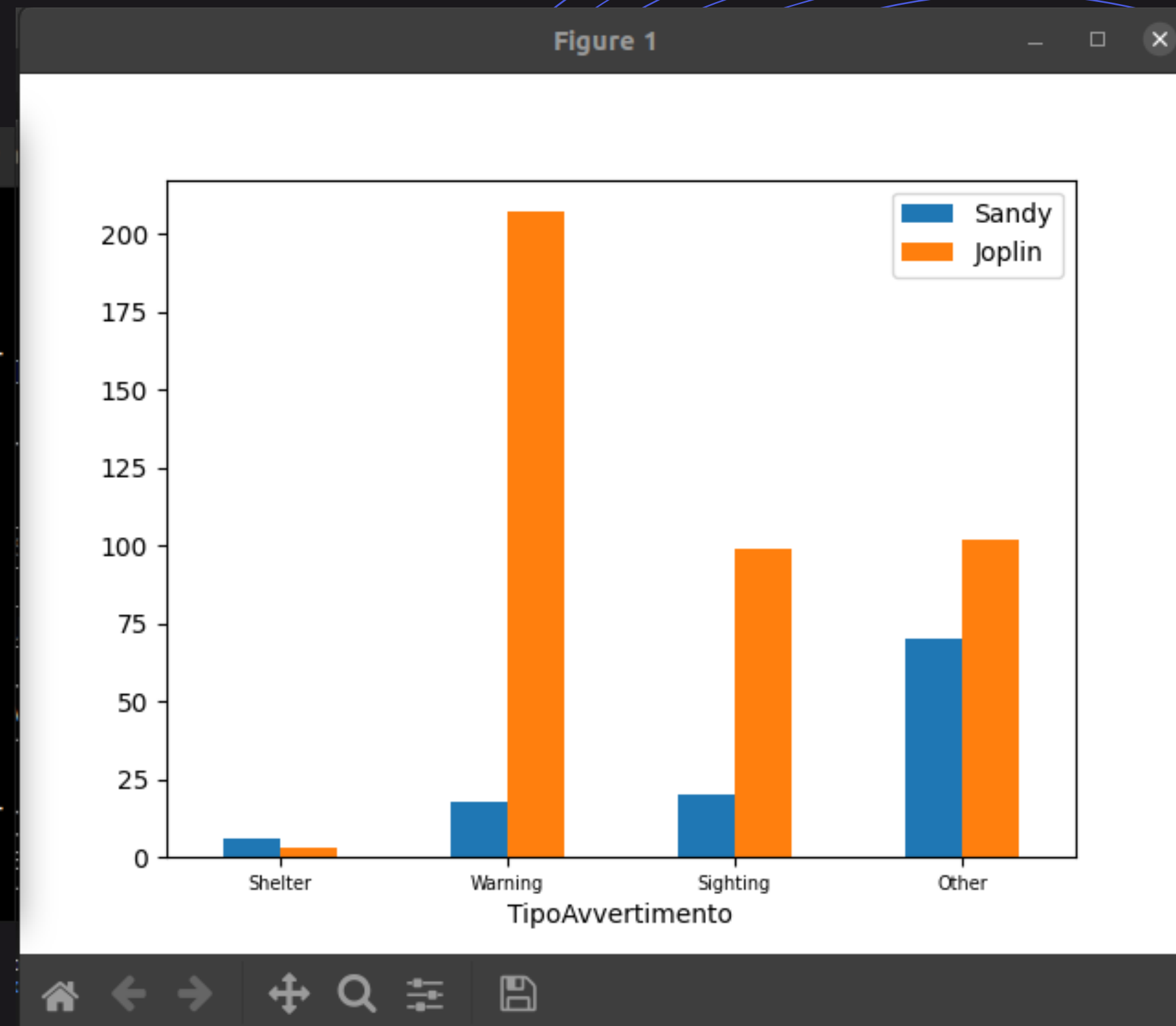
Confronto ==> Numero Tweets

Confronto ==> Tipo di Avvertimenti

Confronto ==> Tipo Di Danni

Confronto ==> Tipo di Media

Torna indietro



ESEMPIO DI ESECUZIONE (3).

12

Sandy Queries

Sandy
Decidi su quali informazioni fare query!

Statistiche ==> Genere di media

Tweet ==> Foto

Tweet ==> Video

Tweet ==> TV

Tweet ==> Web

Tweet ==> Altri media

Torna indietro

Risultato Queries

1) even the rats left town RT @YGNw: Eery images from the empty #NYC subways <http://t.co/eCGqw1ek> #sandy

2) An inside look at NYU hospital's near disaster amid #Sandy by @DrLaPook <http://t.co/DkPjVctS>

3) Photos: The Hurricane Sandy Aftermath: She came she saw she destroyed. <http://t.co/AyEyfY4k> #Sandy #P2 #Weather

4) RT @usefularts: Have you seen @google's storm data tracking map? Adding time lapse for context would be nice <http://t.co/PXvVb0v4> #ddj #data

5) \But... RT @thinkprogress: Bush's FEMA director Scott Brown criticizes Obama for responding to Sandy too "\"quickly\"" <http://t.co/xZgyzZz0>\""

6) The Science of Hurricane Sandy Liveblog <http://t.co/Beub7P3E>

7) \GET THIS NEWS: "\"One of the biggest storms of our lifetimes is unfolding right now.\"" <http://t.co/GKrfn4Qf>\""

8) hurricane #Sandy dashboard for #NYC <http://t.co/F1H6cWCh> created by @melanarchy #graphs #weather

GRAZIE PER
L'ATTENZIONE