

# Report

November 23, 2015

### 0.0.1 Problem Definition

Dato uno stream di tweet (ordinato temporalmente) l'obiettivo è quello di individuare dei "topic" o "event". In letteratura esistono due principali metodologie:

- document-pivot
- feature-pivot

In questo lavoro è stata adottata la prima metodologia, l'obiettivo è quindi suddividere lo stream in cluster tali che ciascun cluster corrisponda a tutti i tweet relativi ad un "evento ". E' altresì necessario, una volta suddiviso lo stream in vari cluster, distinguere quali sono realmente eventi (flashmob) da quelli che non lo sono, per mezzo di un classificatore.

La definizione di evento utilizzata è quella utilizzata per il Topic Detection and Tracking (TDT) :

**Definizione 1.** un evento è qualcosa che accade in un luogo specifico in un tempo specifico

L'intero processo può essere suddiviso in 5 attività come descritto in figura [Allan:2002:TDT:772260] 2002:TDT: