

**Management delle informazioni e  
gestione della conoscenza  
AA 2021-22**

**Text Mining con Rapidminer Studio**

Roberto Boselli  
roberto.boselli@unimib.it

# Rapidminer Studio, introduzione

---



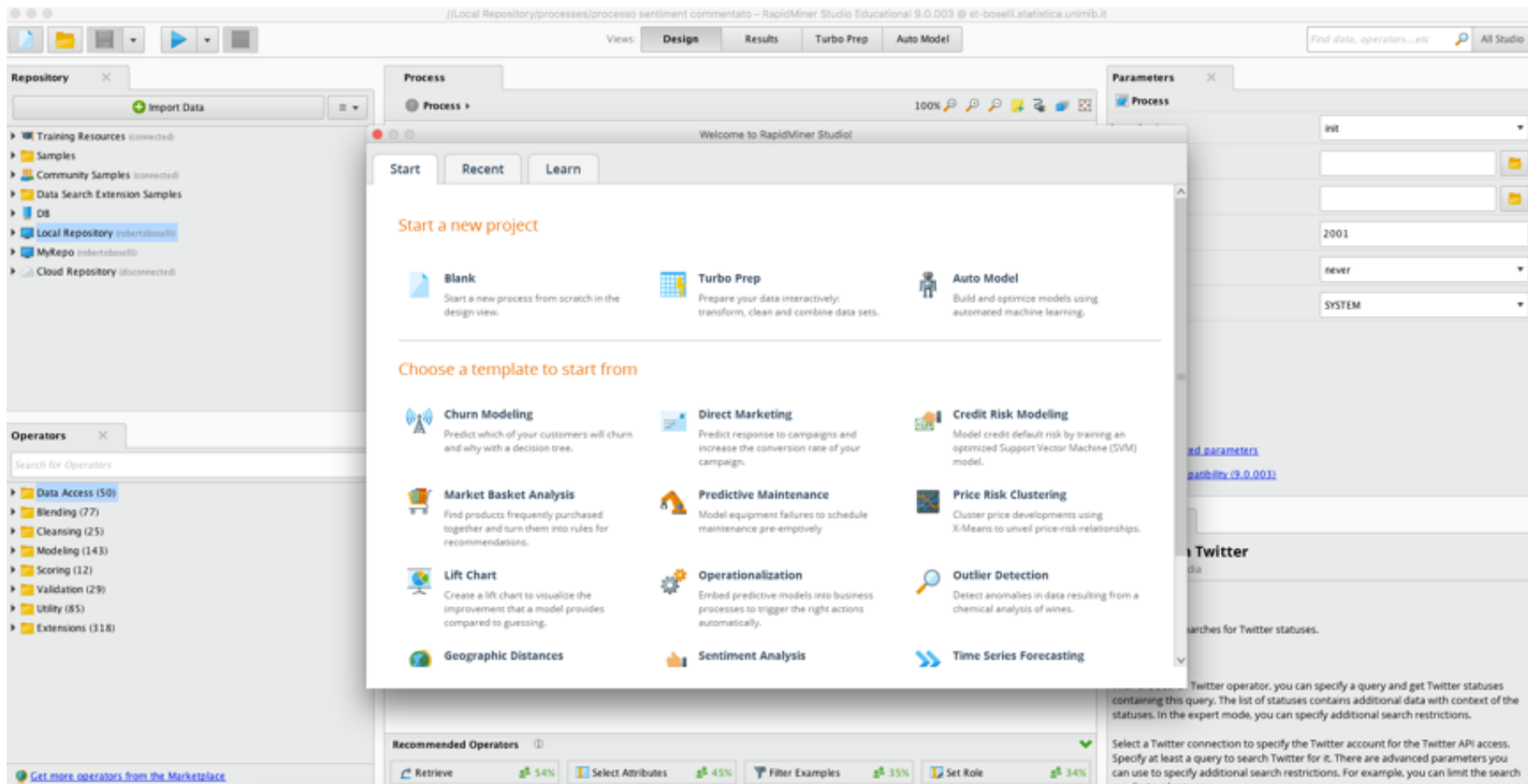
## Installazione e licenza

---

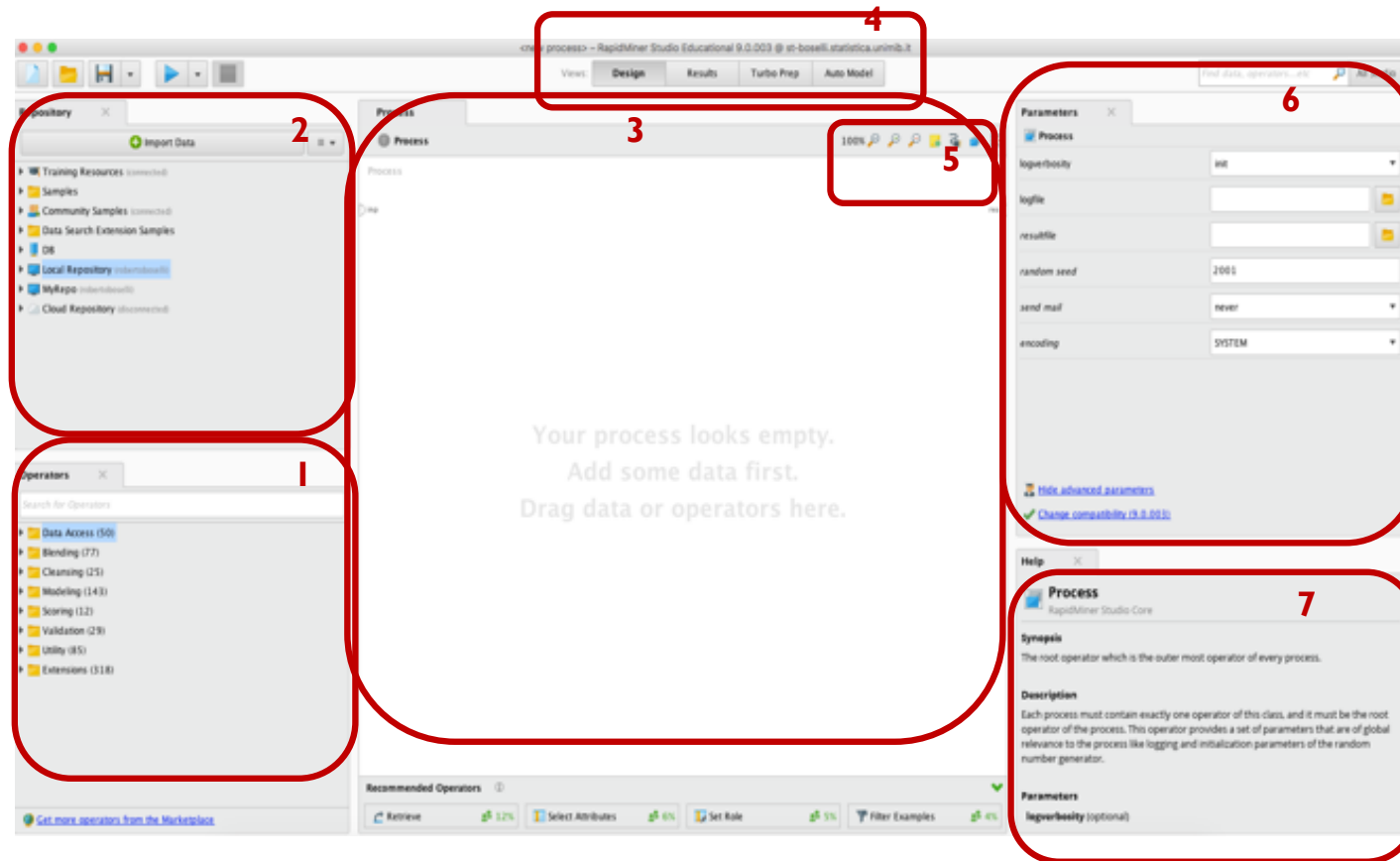
- Scaricare Rapidminer Studio (9.10) dal sito <https://rapidminer.com/get-started/> (è gratuito, versione 30 gg.)
- Dopo averlo installato nel proprio pc, aprire il programma, **creare un account**
- Dalla piattaforma scaricare:
  1. Licenza da inserire per usarlo illimitatamente (fino a luglio 2022)
  2. Guida di installazione Rapidminer, estensioni e connessioni



# Rapidminer Studio 9.10



# New process



**1 Operators** Building blocks used to create RapidMiner processes

**2 Repositories** Storage for data and processes

**3 Process view (Main Process)** Working area for building processes

**4 Views** Buttons for accessing specific functionality




**5 Ports** Input and output mechanisms for operators and processes

**6 Parameters** Settings that modify operator behavior

**7 Help** Context-sensitive help for selected operator.

## Estensioni da installare

---

- Web mining 9.7 
  - Text processing 9.3.1 
  - MeaningCloud 2.1 
- 
- Per installarle: menu Extensions/Marketplace cercare il nome e installare l'estensione una alla volta, alla fine di tutte le installazioni riavviare il programma



# Connessione MeaningCloud API



- 
- ▶ Estensione “MeaningCloud text analytics”
  - ▶ Dalla versione 9.3 di Rapidminer non si deve creare una connessione, ma inserire la Key nelle preferenze del software:
    - ▶ Menu Rapidminer Studio/Preferences: cercare la voce MeaningCloud e incollare la key nel campo vuoto

LINK:

<https://www.meaningcloud.com/developer/>

- ▶ N.B. Nelle versioni precedenti di Rapidminer invece la Key andava incollata come parametro degli operatori di MeaningCloud (es. Sentiment Analysis)



## Scaricare tweets via API, connessione

---

- Per poter scaricare dei tweets e analizzarli in Rapidminer bisogna prima di tutto registrarsi in Twitter (avere un **account Twitter**)
- In Rapidminer cliccare sul menu Connections/Create Connections e aggiungere una nuova connessione di tipo Twitter
- Dare un nome alla nuova connessione (si consiglia: Twitter)
- Access token: cliccare su tasto dx del campo testo
- Richiedere un access token (un codice numerico)
- Permettere a Rapidminer di accedere al vostro account Twitter
- Copiare l'access token (numero) e incollarlo nella finestra della Connection di Rapidminer -> Complete
- Testare la connessione





# Processo di scarico di tweets (hashtag)

---

- ▶ Per scaricare tweets partendo da un hashtag:
  1. Trascinare la connection Twitter nell'area di lavoro (da Local Repository/Connections)
  2. Collegare operatore Search Twitter:
    - Parametro query: scrivere una keyword di ricerca (per es. apple) seguita da `-rt -http -https` (per evitare retweet e links)
    - Limit = 1000 (o numeri >)
    - Language = en (it o altro)
  3. Collegare operatore Write Excel
    - ▶ Dare un nome al file (cambiare nome ogni volta che si scarica)
    - ▶ 4. Collegare porta uscita dell'operatore
    - ▶ 5. Run

