

Specifiche di Progetto

Opencoesione Analysis

Open Coesione Analysis è un progetto di analisi e visualizzazione avanzata dei dati pubblici relativi ai finanziamenti dei progetti territoriali in Italia, con particolare attenzione ai fondi di coesione.

L'obiettivo è fornire strumenti intelligenti per:

- analizzare i ritardi, gli scostamenti finanziari e le criticità dei progetti;
- individuare trend e anomalie;
- supportare l'allocazione più efficiente delle risorse pubbliche.

Attraverso una dashboard interattiva, il progetto offre un'interfaccia intuitiva per cittadini, ricercatori, enti pubblici e decisori politici, basata su tecnologie di big data analytics (Apache Spark) e moderne architetture web (Spring + Angular).

Il progetto integra dati di OpenCoesione con informazioni demografiche ISTAT per arricchire le analisi a livello territoriale e strutturale.

Il secondo obiettivo di questo progetto è mostrare come l'uso di Apache Spark aiuti la manipolazione di Big Data

Dataset

- **Fonte principale:** [Progetti con tracciato esteso – OpenCoesione](#)
 - Altre Fonti:
 - [ISTAT](#)
-

Obiettivi

1. [Analisi dei Ritardi](#)
2. [Analisi Finanziaria](#)
3. [Modellazione Predittiva](#)
4. [Individuazioni fasi critiche](#)
5. [Impatto COVID](#)
6. [Grafo bipartito Beneficiari-Attuatori](#)
7. [Analisi Territoriale](#)
8. [Benchmarking](#)

Analisi dei Ritardi

Identificazione e visualizzazione dei progetti in ritardo, considerando tutte le fasi del ciclo di vita (progettazione, esecuzione, collaudo).

Analisi Finanziaria

Analisi dei progetti in cui si osservano finanziamenti sovrastimati o economie elevate rispetto ai fondi stanziati, per facilitare la ripartizione delle risorse

Modellazione Predittiva

Sviluppo di un modello di classificazione binaria tramite **Spark MLlib**, per prevedere se un progetto subirà un ritardo o meno, sulla base delle caratteristiche strutturali del progetto.

Lo studio sarà suddiviso in tre fasi temporali:

- **Pre-COVID**
 - **Durante COVID (2020–2021)**
 - **Post-COVID**
-

Individuazioni fasi critiche

Identificazione delle **fasi più critiche** nel ciclo di vita del progetto tramite:

- Calcolo della **durata media per fase**
- Calcolo della **deviazione standard per fase** (indicatore di instabilità temporale)

Questi valori permetteranno di individuare quali fasi necessitano attenzione in termini di gestione.

Impatto COVID

Analisi comparativa tra i progetti attivati prima, durante e dopo il periodo pandemico (2020–2021), per misurare l’impatto su:

- Durata complessiva
 - Ritardi
 - Fondi erogati e economie
-

Grafo bipartito Beneficiari-Attuatori

Costruzione di un grafo bipartito con:

- Nodi: **Beneficiari** (DENOMINAZIONE_BENEFICIARIO) e **Attuatori** (DENOMINAZIONE_SOGGETTO_ATTUATORE)
- Archi: relazioni progetto tra i due

L’analisi tramite **GraphX (Spark)** permetterà di:

- Individuare i **nodi centrali**
 - Riconoscere **community** (gruppi di soggetti che collaborano frequentemente)
-

Analisi Territoriale

Confronto geografico tra Nord, Centro e Sud per:

- Valore totale dei fondi ricevuti
 - Frequenza e intensità dei ritardi
 - Entità delle economie pubbliche realizzate
-

Benchmarking

Tecnologie confrontate

- Apache Spark
- PostgreSQL
- MongoDB

Metriche di valutazione

- Tempo medio di query

- Efficienza di lettura/scrittura su:
 - **HDFS**
 - **File system locale**
 - Performance al crescere del volume dati
 - Gestione della cache e utilizzo della memoria
-

Dashboard e Visualizzazioni

Le analisi verranno presentate tramite una **dashboard interattiva** con:

- Grafici geografici e temporali
 - Indicatori aggregati
 - Funzionalità di filtro per categoria, territorio, soggetto
-

Colonne e Variabili di Interesse

Trasversalmente a tutte le analisi, si farà uso delle seguenti sezioni del dataset:

- **Territorio:** Regione, Provincia, Comune
 - **Soggetti:** Denominazione e forma giuridica di attuatori e beneficiari
 - **Classificatori:** Settore, Sottosettore, Categoria, Tipologia, Natura
 - **Finanza:**
 - TOT_PAGAMENTI
 - ECONOMIE_TOTALI_PUBBLICHE
 - OC_FINANZ_TOT_PUB_NETTO
 - FINANZ_TOTALE_PUBBLICO
 - Dati demografici
-

Architettura Tecnologica

- **Frontend:** Angular
- **Backend:** Spring Boot con API REST
- **Elaborazione dati:** Apache Spark (SparkSQL, MLlib, GraphX)
- **Orchestrazione:** Kubernetes per gestione dinamica dei job Spark in ambienti distribuiti
- **Persistenza:** PostgreSQL, MongoDB, FS locale e HDFS
- **Visualizzazione:** Dashboard Angular

