

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE

Corso di Laurea in
INFORMATICA PER IL MANAGEMENT

**LA DIMENSIONE DISCORSIVA
DELLA POLITICA DI COESIONE.
CONFRONTO FRA
*CONTENT ANALYSIS E TOPIC MODELING***

Tesi di Laurea in
Strategia Aziendale

Relatore

Chiar.mo Prof. Edoardo Mollona

Correlatore

Prof. Luca Pareschi

Presentata da

Silvia Di Fiore

Sessione II

Anno Accademico 2017/2018

INDICE

INTRODUZIONE	I
CAPITOLO I	
<i>TEXT MINING</i>	1
1.1 Cenni Storici	1
1.2 Dall'informazione al <i>Dato Testuale</i>	3
1.2.1 <i>Preparazione del testo</i>	3
1.2.2 <i>Preprocessing</i>	4
1.2.3 <i>Creazione delle Feature</i>	6
1.3 Fase di Modellazione: Tecniche di <i>Text Mining</i>	7
1.4 Analisi dei risultati	8
CAPITOLO II	
LA POLITICA DI COESIONE DELL'UE	9
2.1 Perché attuare una politica di coesione in Europa?	10
2.2 Tratti storici importanti della politica regionale Europea	11
2.3 Fondi Strutturali Europei	21
2.4 L'evoluzione del bilancio nel corso degli anni	24
2.5 L'impatto dei Fondi di Coesione in Italia	26
CAPITOLO III	
METODOLOGIE: <i>CONTENT ANALYSIS</i>	31
3.1 Definizione e cenni storici della <i>Content Analysis</i>	32
3.2 Fasi della <i>Content Analysis</i>	34
3.3 Limiti e Osservazioni	39

CAPITOLO IV	
TOPIC MODELING	41
4.1 Algoritmo <i>Latent Dirichlet Allocation</i>	42
4.2 Software <i>Mallet</i>	44
CAPITOLO V	
CASO DI STUDIO	49
5.1 Analisi Qualitativa	50
5.1.1 Risultati analisi qualitativa	52
5.2 Analisi attraverso il <i>Topic Modeling</i>	60
5.2.1 Risultati analisi a 10 Topic	64
5.2.2 Risultati analisi a 15 Topic	68
5.3 Confronto tra i risultati ottenuti con l'analisi <i>Topic Modeling</i>	74
5.4 Confronto dei risultati tra <i>Analisi Qualitativa</i> e <i>Topic Modeling</i>	76
CONCLUSIONI	83
BIBLIOGRAFIA	85
SITOGRAFIA	90

INTRODUZIONE

Il presente lavoro consiste nell'analizzare una raccolta di articoli, digitali e non, attraverso due tipologie di analisi: *Analisi Qualitativa*, e *Analisi Topic Modeling*.

Lo scopo primario è quello di valutare e confrontare tali modelli, applicandoli ad una raccolta di articoli di un Quotidiano che trattano di un argomento specifico.

La prima parte della tesi si concentra sul concetto del *Text Mining*. Viene fornita la sua definizione, i possibili campi di utilizzo, i suoi processi di sviluppo nel tempo e, infine, vengono espone le diverse tecniche di *Text Mining*.

Il capitolo seguente tratta l'argomento scelto su cui applicare le due tipologie di analisi, la *Politica di Coesione dell'Unione Europea*. In questo capitolo descriveremo il motivo per il quale è stata attuata una Politica di Coesione in Europa ripercorrendo i tratti storici più salienti.

Elencheremo, inoltre, i vari fondi Strutturali messi a disposizione dall'*Unione Europea* per eliminare il divario tra le varie regioni e infine analizzeremo i risultati ottenuti nell'arco degli anni (dagli anni '60 a oggi), in termini di Bilancio Europeo e l'impatto che hanno avuto sul nostro territorio Italiano.

Il terzo capitolo inizia con una breve ricostruzione delle radici storico-concettuali del metodo di ricerca, denominato *Content Analysis* (Analisi del Contenuto), usato nelle Scienze Sociali per lo studio di materiale testuale (sia esso scritto, verbale, iconico o audiovisivo). Ne descriveremo, inoltre i metodi, le tecniche e i limiti.

Il quarto capitolo entra nel merito dell'*Analisi Topic Modeling*, ovvero un modello probabilistico il cui scopo è rivelare i temi sottostanti ad una collezione di documenti.

Nello specifico abbiamo descritto l'algoritmo *Topic Model* più conosciuto e utilizzato, il *Latent Dirichlet Allocation (LDA)* e un *software* specifico che lo utilizza, *Mallet*.

La fase finale dell'elaborato guida il lettore attraverso un esempio pratico, in cui i modelli precedentemente descritti vengono applicati ad articoli riguardanti la *Politica di Coesione*.

Vengono elencati nel dettaglio i risultati ottenuti dalle due analisi, attraverso apposite tabelle e grafici, e infine confrontati.

L'ultimo passaggio è dedicato alle conclusioni, dove vengono discussi i vantaggi e gli svantaggi di entrambi i modelli e alcune considerazioni in merito.

CAPITOLO I

TEXT MINING

Il *Text Mining* è un'area relativamente nuova dell'informatica e il suo utilizzo cresce in modo esponenziale dacché i dati non strutturati disponibili continuano ad aumentare sia in termini di rilevanza che di quantità. Il concetto che sta alla base del *Text Mining* è quello di riuscire ad estrarre pochissime informazioni rilevanti dall'immensa quantità di dati a disposizione senza doverli elaborare tutti.

In questo primo capitolo verranno trattati gli aspetti principali del *Text Mining* dando, inoltre, un breve accenno su come è nata questa tecnica ripercorrendo i vari studi effettuati.

Verranno risposte quattro principali domande che permetteranno al lettore di comprendere a pieno l'argomento, ovvero: Cosa si intende per *Text Mining*; come funziona il processo del *Text Mining*; quali e come vengono applicate le tecniche di *Text Mining*; come vengono analizzati i risultati ottenuti.

1.1 Cenni Storici

Gli studi basati su questa tecnica iniziano negli anni '60, per poi svilupparsi lentamente nel tempo e raggiungere il picco massimo in anni recenti, grazie all'aumento delle quantità di dati in formato digitale. Inizialmente lo studio era focalizzato sull'indicizzazione e sulla ricerca dei documenti.¹

¹ MARON, M., KUHNS, J., 1960. On relevance, probabilistic indexing and information retrieval. In *Journal of the ACM*, Vol. 7, Pp. 216-244.

Nel primo caso il problema era incentrato sulla classificazione, ovvero, dato un insieme di categorie, ogni categoria veniva automaticamente assegnata ad uno specifico documento.

Nel secondo caso, invece, il problema era incentrato sulla ricerca di documenti che trattano un determinato argomento all'interno di una vasta mole di dati.²

Con il passare degli anni si inizia a diffondere l'idea di una descrizione matematica dei testi scritti. Questa, permette di compiere i primi studi sull'analisi *cluster* applicata ai documenti, volta alla selezione e al raggruppamento di elementi omogenei in un insieme di dati.³

Le *Tecniche di Clustering* si basano su misure relative alla somiglianza tra gli elementi. Si affermano quindi diverse definizioni di similarità tra testi scritti, che permettono l'implementazione dei principali algoritmi di *Data Mining*. Uno dei maggiori problemi degli anni '60-'70 era la scarsa disponibilità di dati e la codifica digitale dei testi. Da qui nasce la necessità di sviluppare efficaci motori di ricerca, sistemi per ricercare le informazioni principali di un dato documento e filtri per mettere in risalto documenti rilevanti.

Successivamente, l'esponenziale crescita del web e dei suoi utilizzatori verificatosi negli ultimi anni hanno contribuito ad incrementare l'aggiunta di nuove fonti di conoscenza creando un gigantesco patrimonio informativo.

Sempre più persone sono state attratte da questo studio in quanto si crede che il suo potenziale commerciale sia superiore a quello dell'estrazione dati. In effetti, basta pensare che la maggior parte della produzione d'informazioni di una società proviene da chat, e-mail, blog e forum.

Con tali prospettive di sviluppo si riesce a dare maggiore efficacia e robustezza all'analisi automatica dei documenti solo grazie al *Text Mining* perché è l'unica tecnologia che esegue uno studio approfondito del "significato" del testo.

² LUHN, Hans Peter, 1959. Auto-encoding of documents for information retrieval systems. In BOAZ, Martha, (Editor), *Modern Trends in Documentation*. London, Pergamon Press. Pp. 45–58.

³ JARDINE, Nick, VAN RIJSBERGEN, Cornelis Joost, 1971. The use of hierarchical clustering in information retrieval. In *Information Storage and Retrieval*, vol. 7, Pp. 217–240

1.2 Dall'informazione al *Dato Testuale*

Oggi giorno, la rete è un “archivio vivente” che contiene una massa ingente di informazione. L'informazione abbandonata a sé stessa rimane offuscata da un apparente disordine. Applicando alcune metodologie di analisi automatica dei testi possiamo imparare a gestire meglio e in modo più proficuo il patrimonio di dati, per ora nascosti, che sono già a nostra disposizione selezionando ciò che è di nostro interesse e visualizzandone sinteticamente i risultati. Prima di procedere con l'illustrazione di questa tecnica dobbiamo però impadronirci delle nozioni fondamentali che riguardano la preparazione dei testi.

Nella rete troviamo un po' di tutto: testi poetici e letterari, articoli di giornali, documenti legislativi, articoli scientifici, pagine web, articoli di Wikipedia, messaggi dei social network. Non vi sono limiti alla varietà di testi che possiamo trovare in Internet.

Il nostro interesse verso l'analisi di uno o più testi deve essere guidato da qualche domanda che vogliamo rivolgere ad essi e da una o più risposte provvisorie che rappresentano le nostre ipotesi di lavoro. In tal senso va selezionata una raccolta coerente di testi, detta corpus, omogenea sotto un qualche punto di vista che verrà successivamente preparato per l'analisi.

1.2.1 Preparazione del testo

La fase di preparazione è essenziale per una corretta scansione del testo secondo l'unità di analisi prescelta. I documenti testuali caratterizzano la maggior parte delle informazioni degli esperimenti sul *Text Mining*, quindi è necessario trasformarli dal formato destrutturizzato in cui si trovano ad un formato strutturizzato adatto per poter applicare gli algoritmi di *data-mining*.

La struttura del testo è molto importante, ad esempio, la stringa di caratteri che compone il titolo sarà determinante per individuare l'argomento trattato dal testo,

oppure, le parole evidenziate in grassetto possono indicare concetti chiave, così come spesso possono essere utili le informazioni riguardanti la data e l'autore *etc.* Questa fase si occupa della pulizia dei formati di gestione del testo e della normalizzazione del testo che consiste nell'uniformare spazi, apostrofi e accenti, riconoscere a priori entità particolari (date, numeri, valute, titoli, sigle, abbreviazioni), nonché nomi, Società, personaggi o espressioni e locuzioni d'interesse.

Fra i tanti formati disponibili, il più diffuso utilizzato nelle tecniche di *Text Mining* è il linguaggio **XML** (*Xtensible Markup Language*) basato sulla suddivisione del testo in campi che danno informazioni sul contenuto (ad esempio titolo, autore, sommario) tramite *tags* (parole chiave).

Solitamente è utilizzato questo linguaggio perché quasi tutti gli *editor* permettono la conversione e l'esportazione automatica di un qualsiasi testo in tale linguaggio.

1.2.2 Preprocessing

Oltre alla preparazione e conversione dei documenti in un formato strutturalizzato, possono essere necessarie varie fasi di rielaborazione dei testi, che coinvolgono strumenti linguistici e di *Natural Language Processing*⁴ a volte molto sofisticati. Non tutte queste fasi sono però necessarie ad un'analisi di *Text Mining*, al contrario, alcune potrebbero fornire informazioni inutili a seconda del tipo di analisi.

Le fasi principali di *preprocessing* sono le seguenti:⁵

1) TOKENIZATION: permette di rompere una sequenza di caratteri in parole/frasi detti "token". I *Token* sono separati l'uno dall'altro da appositi caratteri, i separatori, che

⁴ Elaborazione del Linguaggio Naturale detta anche **NLP** (*Natural Language Processing*) è il processo di trattamento automatico mediante un calcolatore elettronico, delle informazioni scritte o parlate in una lingua naturale.

⁵ DULLI, S., POLPETTINI, P., TROTTA, M., (a cura di), 2004. *Text mining: teoria e applicazioni*. Milano, Franco Angeli Editore.

possono essere spazi bianchi, tabulazioni, terminatori di paragrafo o elementi di punteggiatura.

2) STEMMING: è basato sul conteggio di quante volte una determinata parola compare nel documento. Abbiamo due tipi di *Stemming*:

a. *Inflectional stemming o lemmatizzazione:* riguarda la normalizzazione di declinazione e coniugazione delle parole, ovvero raggruppa le varie forme flesse di una parola in modo che possano essere analizzate come una sola entità. La forma normalizzata è detta lemma;

b. *Root stemming:* è un procedimento che va più in profondità, mirato ad eliminare prefissi e suffissi delle parole. In questo caso si risale non solo da forme diverse (vado, andiamo) ad un'unica parola (il lemma andare), ma anche da forme diverse di lemmi diversi (andiamo, andamento) ad un'unica radice (and-), detta anche morfema lessicale. In questo modo, alla fine della fase di *Stemming* resteranno molti meno termini diversi, mettendo meglio in evidenza le informazioni riguardanti i temi trattati dal documento.

3) SEPARAZIONE DI FRASI: Come la Procedura di Tokenization prevede la divisione di una stringa di caratteri in token, questa fase prevede la suddivisione del testo in frasi.

4) PART OF SPEECH TAGGING: Dopo aver diviso il testo in frasi, si può procedere con un'assegnazione del ruolo grammaticale, attribuendo ad ogni Token il ruolo di nome, verbo, aggettivo, avverbio ecc. Il riconoscimento del ruolo grammaticale di un termine si basa in gran parte sulla costruzione (manuale) di dizionari che riportano un elenco di termini associati ai ruoli.

5) RICONOSCIMENTO DI SINTAGMI: consiste nel riconoscimento di sintagmi elementari. In modo semplicistico, possiamo definire il sintagma come un elemento

della frase che rappresenta un'unità di significato. Questa fase è perciò utile sia come primo step per l'analisi del periodo, che per individuare relazioni semplici tra elementi nel testo, espresse da frasi elementari.

6) NAMED ENTITY RECOGNITION: sono degli elementi noti che ci permettono di acquisire molte informazioni. Solitamente sono nomi di luoghi, di aziende, espressioni temporali come date ed ore e quantità interessanti. Anche in questo caso, il lavoro principale consiste nella redazione di dizionari specifici, che riportino tutti i termini di interesse, ma anche nella costruzione di specifici metodi di machine Learning.

7) PARSING: consiste nell'analisi delle relazioni tra i Token di una frase, e nell'assegnazione di un ruolo semantico. L'*Output del Parsing* è solitamente un albero, dove nodi rappresentano le relazioni e le foglie i Token. Con tale albero è possibile ricostruire buona parte dell'analisi logica di una frase ed estrarne informazioni approfondite. In tal modo, si possono costruire sistemi in grado di riconoscere nel testo fatti specifici.

1.2.3 Creazione delle Feature

Una volta effettuata la *Fase di Preprocessing*, per il nostro testo si ottiene un *Vettore di Feature* che lo caratterizza. La creazione di *Vettori di Feature* a partire dai documenti si può fare in diversi modi. Essa dipende principalmente dal tipo di analisi che si vuole operare sui documenti e dalle fasi di *Preprocessing* eseguite.

Per analizzare l'argomento trattato dal documento, ci focalizziamo sui termini usati.

Le Fasi di *Preprocessing* necessarie sono perciò quella di *Tokenization* e *Stemming*. Un *Vettore di Feature* è dunque costruito in modo che ogni componente corrisponda ad un *Token*.

In alcuni progetti di *Text Mining* vengono selezionati solo le *Feature* più rilevanti che caratterizzano il testo prima di passare alla *Fase di Modellazione*.

1.3 Fase di Modellazione: Tecniche di *Text Mining*

Sui dati adesso strutturati, andremo ad applicare tecniche di *Text Mining* che ci permetteranno trovare le informazioni tematiche nascoste in un testo, per facilitare il processo di archiviazione e di costruzione di una mappa logica della conoscenza.

Queste tecniche si basano su determinati algoritmi che selezionano le parti rilevanti di un documento ed eliminano quelle non rilevanti.

Le tecniche più diffuse sono:

- **Classificazione:** tecnica che permette di assegnare testi a categorie predefinite, basandosi sul loro contenuto.
- **Clustering:** tecnica non supervisionata che permette di raggruppare gruppi di documenti con caratteristiche simili tra loro in “clusters”. In un cluster quindi troveremo una serie di testi raggruppati che si riferiscono ad uno stesso argomento o parole chiavi identiche.
- **Estrazione di Termini:** tecnica che estrae parole o concetti, in base alla sua categoria (ad esempio, nomi di aziende).
- **Estrazione di Relazioni:** tecnica utilizzata per estrarre la relazione contenuta nel testo (ad esempio, Rome-Capital-Italy).
- **Estrazione di Segnali Deboli:** tecnica utilizzata per estrarre espressioni (ad esempio, la frase “Agnelli sposa la General Motors” è riconosciuta come “partnership tra FIAT e GM”).
- **Identificazione di Sigle:** tecnica che permette il riconoscimento di sigle.

- **Riconoscimento di Sinonimi:** tecnica che permette di riconoscere ad esempio “casa” ed “abitazione” come sinonimi.
- **Riconoscimento di Polisemi:** tecnica che permette di riconoscere i diversi significati di una parola.
- **Linguaggi Specifici:** tecnica che permette il riconoscimento immediato di abbreviazioni, sinonimi e polisemi, individuando la lingua specializzata relativa al documento in oggetto utilizzando dizionari.

1.4 Analisi dei risultati

Dopo aver effettuato la fase di modellazione, i risultati ottenuti vengono valutati e analizzati sulla base di diversi parametri che variano in base alla tecnica di modellazione utilizzata. Se si sfrutta la classificazione la valutazione avverrà sulla base dell'accuratezza, del richiamo e della precisione.

Nel caso in cui i risultati non siano soddisfacenti si riparte da una fase precedente nel processo a spirale in maniera tale da apportare modifiche alle scelte fatte, nell'intento di migliorare.

CAPITOLO II

LA POLITICA DI COESIONE DELL'UNIONE EUROPEA



Fig. 1 Logo per la Politica di Coesione Europea.
Fonte: EUROPEAN COMMISSION.EU, < www.ec.europa.eu/>.

In questo secondo capitolo tratteremo gli aspetti principali della *Politica di Coesione* dell'Unione Europea, offrendo una risposta ai quesiti seguenti: a cosa serve, perché, quali sono le sue prospettive e con quali mezzi cerca di raggiungerle?

Prima di cominciare però, bisogna fare chiarezza sul significato dei termini *Politica Regionale* e *Politica di Coesione* in quanto vengono spesso scambiati come sinonimi. La differenza sta nel fatto che la *Politica di Coesione* è finanziata dal **FESR** (*Fondo Europeo di Sviluppo Regionale*), dal **FSE** (*Fondo Sociale Europeo*) e dal *Fondo di Coesione*, e perciò è un concetto più ampio della *Politica Regionale*, la quale è collegata specificatamente alle attività del FESR.

Ci riferiremo quindi a quella di *Coesione* come la *Politica unitaria ed organica* che scaturisce dalla riforma dei fondi strutturali attuata nel 1988 dalla *Commissione* presieduta allora dal politico ed economista francese Jacques Delors.

Il presente capitolo vuole affrontare in primo luogo le ragioni di fondo dell'esistenza di questa *Politica Europea*.

2.1 Perché attuare una Politica di Coesione in Europa?

Notevoli sono le disparità che esistono tra le regioni dell'Unione Europea per quanto riguarda soprattutto il prodotto interno lordo, il reddito e i livelli di occupazione.

Le regioni più ricche rispetto alla media UE sono principalmente concentrate in Gran Bretagna, in Olanda, in Belgio, nel sudovest della Germania, nell'Austria occidentale e nell'Italia del nord, mentre la parte sud est del territorio europeo, ovvero Italia meridionale, Grecia e del PECO (Paesi dell'Europa Centrale e Orientale), registra un livello di PIL al di sotto della media. E' appunto la presenza di queste disparità economiche e sociali interregionali che ha portato l'Unione Europea ad attuare una politica di coesione con lo specifico scopo di ridurle.

Il libro verde sulla coesione territoriale presentato dalla Commissione a Consiglio e Parlamento Europeo il 6 ottobre 2008 si apre in questo modo:

*Dalla tundra ghiacciata del circolo artico alle foreste tropicali della Guyana, dalle Alpi alle isole della Grecia, dalle città globali di Londra e Parigi alle cittadine e ai villaggi vecchi di secoli, l'UE racchiude una diversità territoriale straordinariamente ricca.*⁶

L'obiettivo della politica di coesione territoriale è quello di garantire lo sviluppo di tutte le regioni europee attraverso la trasformazione di queste diversità in un punto di forza per la crescita sostenibile e a lungo termine di tutta l'Unione aiutando ogni regione a realizzare le proprie potenzialità, migliorare la competitività e l'occupazione investendo nei settori con un elevato potenziale di crescita e innalzare il tenore di vita nei vari paesi.

⁶ COMMISSIONE DELLE COMUNITÀ EUROPEE, 2008. *Libro verde sulla coesione territoriale. Fare della diversità territoriale un punto di forza*. Bruxelles, (6.10.2008), <http://ec.europa.eu/regional_policy/archive/consultation/terco/paper_terco_it.pdf>, (Ultimo Accesso, 12-06-2018), P. 3.

In un'economia globalizzata fortemente competitiva diventa essenziale riuscire a sfruttare nel modo migliore le proprie risorse specifiche. Una politica pubblica può aiutare anche i territori più piccoli e lontani mettendo a disposizione fondi che sarebbe altrimenti molto difficile rinvenire sul territorio.

Di fondamentale importanza diventa anche la programmazione degli interventi in modo da elaborare una risposta unitaria ai problemi comuni, oltrepassando le frontiere determinate dalle divisioni amministrative e accrescendo così i profitti grazie alla cooperazione e alla creazione di sinergie.

I fondi strutturali sono stati progettati a favore di tutte le regioni svantaggiate con il fine di permettere lo sviluppo dei settori in cui possiedono un vantaggio comparato e aiutare il superamento dei fattori che limitano la crescita. La sfida più importante consiste nell'assicurare uno sviluppo territoriale equilibrato e sostenibile di tutta l'Unione, rafforzando la competitività di tutte le regioni e garantendo allo stesso tempo la coesione sociale

2.2 Tratti storici importanti della Politica Regionale Europea

Nel 1957, con il *Trattato di Roma*, inizia la storia della politica regionale Europea, nello stesso momento in cui nascono le Comunità Europee.

Nei primi anni le risorse destinate a questo scopo non erano molte, ma con il passare del tempo la loro importanza è progressivamente cresciuta, arrivando negli ultimi anni a rappresentare più di un terzo dell'intero bilancio dell'Unione.

Verrà mostrata in seguito una tabella contenente in maniera schematica e cronologica i tratti storici importanti nella storia della *Politica Regionale Europea*

PERIODO	EVENTO
25/03/1957	TRATTATO DI ROMA
1958	Vengono istituiti due fondi settoriali: il Fondo sociale europeo (FSE) e il Fondo europeo agricolo di orientamento e di garanzia (FEAOG).
04/04/1962	Regolamento (CEE) del consiglio n. 25/1962: istituisce il fondo europeo agricolo di orientamento e garanzia (FEAOG).
18/03/1975	Regolamento (CEE) del consiglio n. 724/1975: istituisce il fondo europeo di sviluppo regionale (FESR).
1981	Ingresso della Grecia nella CEE.
1985	La commissione presenta un libro bianco dal titolo "Il completamento del mercato interno".
01/01/1986	Ingresso di Spagna e Portogallo nella CEE: aumentano ulteriormente gli squilibri regionali.
17/02/1986	Firmato l'atto unico europeo: riconosce formalmente tra gli obiettivi la coesione economica e sociale.
1987	Presentato al parlamento europeo il pacchetto Delors I.
1988	Vengono introdotti cinque obiettivi prioritari e i principi di concentrazione, programmazione, addizionalità e partenariato.
07/02/1992	FIRMA DEL TRATTATO DI MAASTRICHT (TUE): • prevista la creazione di un fondo di coesione e due organi consultivi, il comitato economico e sociale e il comitato delle regioni.
1992	Il Consiglio raddoppia le risorse annue per i fondi strutturali per il periodo 1994-1999.
1995	In seguito all'adesione di Svezia e Finlandia viene introdotto un sesto obiettivo della politica regionale riguardante le regioni a bassissima densità abitativa.
1999	Gli obiettivi della politica regionale scendono da sei a tre.
01/05/2004	Dieci nuovi stati aderiscono all'UE (Cipro, Estonia, Lettonia, Lituania, Malta, Polonia, Repubblica Ceca, Slovacchia, Slovenia e Ungheria) e le disparità regionali si allargano ancora di più.
2006	Nuova riforma della politica regionale con nuovi obiettivi.
01/01/2007	Romania e Bulgaria entrano nell'unione, a cui ora appartengono 27 paesi.

Tab. 1 Vengono riportati in maniera schematica i principali avvenimenti storici che hanno caratterizzato la formazione della Politica Regionale Europea dal 1957, 2007.

Fonte: Ricostruzione personale.

Il punto cardine delle politiche di coesione fin dal *Trattato di Roma* è la necessità di garantire uno sviluppo adeguato e armonioso a tutte le regioni dell'Unione Europea. I paesi membri fondatori, ovvero *Francia, Germania, Italia, Olanda, Belgio e Lussemburgo*, nel preambolo inseriscono l'obiettivo di *rafforzare l'unità delle loro economie e di assicurarne lo sviluppo armonioso riducendo le disparità fra le differenti regioni e il ritardo di quelle meno favorite*.⁷

Con tale Trattato viene istituito anche il primo dei fondi strutturali, il *Fondo Sociale Europeo (FSE)*,⁸ il cui scopo è migliorare l'occupazione e le possibilità di impiego. Nel 1962 viene creato il *Fondo Europeo Agricolo di Orientamento e Garanzia (FEAOG)*.⁹

In questo primo periodo non esistono ancora forme dirette di aiuto alle regioni svantaggiate, ma gli interventi restano concentrati soprattutto a livello nazionale con il finanziamento di progetti predefiniti dagli Stati membri. Solo nel 1975 arriva la creazione di un *Fondo Europeo di Sviluppo Regionale (FESR)* che mira a ridurre i problemi economici, ambientali e sociali che affliggono le aree urbane, investendo principalmente nello sviluppo urbano sostenibile. La vera e propria rivoluzione ha inizio negli anni '80, infatti, nel giugno del 1985 la *Commissione Europea* inizia a stimolare il dibattito sul tema con il libro bianco presentato al *Consiglio Europeo* di Milano dal titolo: "Il completamento del mercato interno", nel quale auspica il raggiungimento del mercato unico entro il 1992.

Nel frattempo l'adesione della *Grecia* nel 1981, della *Spagna* e del *Portogallo* nel 1986, provocano un ulteriore aumento delle disparità tra le regioni della CEE.

Il 17 Febbraio 1986 venne firmato a *Lussemburgo* l'*Atto Unico Europeo (AUE)*, al fine di rilanciare l'Integrazione Europea e portare al termine la realizzazione del mercato interno.

⁷CEE (*Comunità Economica Europea*), 1957. *Trattato che istituisce la Comunità economica europea (CEE). Preambolo*, ROMA, (25/3/1957), pag. 11.

⁸ Istituito dall'art. 123 del *Trattato di Roma* del 1957 e disciplinato con regolamento CEE del Consiglio n. 9/1960 del 25 agosto 1960.

⁹ Istituito dal regolamento CEE del Consiglio n. 25/1962 del 4 aprile 1962.

Tale atto riconosce formalmente tra gli obiettivi prioritari della Comunità la coesione economica e sociale, tramite il coordinamento delle politiche economiche nazionali e l'attuazione di politiche comuni.

All'interno del documento le disparità regionali vengono definite un elemento di freno per la realizzazione del mercato interno Europeo e per l'approfondimento dell'integrazione economica stessa. Per favorire quest'ultima occorreva intensificare gli sforzi finanziari destinati alla coesione economica e sociale.

Il *Consiglio Europeo*, nel marzo 1988, approva il cosiddetto pacchetto "Delors I", che contiene una serie di norme volte a riordinare e riorganizzare radicalmente la finanza comunitaria, riguardanti la necessità di dotare la Comunità di risorse adeguate per attuare le riforme indispensabili alla concretizzazione dell'Atto unico.¹⁰

Vengono emanati cinque nuovi regolamenti: il Regolamento Quadro n. 2052/88,¹¹ che enunciava gli obiettivi dei fondi e definiva i principi base del loro funzionamento; un Regolamento Di Coordinamento, il n. 4253,¹² che prevedeva una gestione maggiormente integrata delle risorse e un regolamento di applicazione per ognuno dei tre fondi: 4254/88/CEE (FESR), 4255/88/CEE (FSE), 4256/88/CEE (FEAOG).¹³

¹⁰VAROTTO, M., 2010. *Le opportunità dell'Unione Europea per gli enti locali*. Milano, Ipsoa Edizioni.

¹¹ Regolamento (CEE) n. 2052/88 del Consiglio, del 24 giugno 1988, relativo alle missioni dei Fondi a finalità strutturali, alla loro efficacia e al coordinamento dei loro interventi e di quelli della Banca europea per gli investimenti degli altri strumenti finanziari esistenti. In *Gazzetta ufficiale delle Comunità europee*. L 185, 15 luglio 1988, (31° anno).

¹² Regolamento (CEE) n. 4253/88 del Consiglio, del 19 dicembre 1988, recante disposizioni di applicazione del regolamento (CEE) n. 2052/88 per quanto riguarda il coordinamento tra gli interventi dei vari Fondi strutturali, da un lato, e tra tali interventi e quelli della Banca europea per gli investimenti e degli altri strumenti finanziari esistenti, dell'altro. In *Gazzetta ufficiale delle Comunità europee*. L 374, 31 dicembre 1988 (31° anno).

¹³ Regolamento (CEE) n. 4254/88 del Consiglio del 19 dicembre 1988 recante disposizioni di applicazione del regolamento (CEE) n. 2052/88 per quanto riguarda il Fondo europeo di sviluppo regionale; Regolamento (CEE) n. 4255/88 del Consiglio del 19 dicembre 1988 recante disposizioni d'applicazione del regolamento (CEE) n. 2052/88 per quanto riguarda il Fondo sociale europeo; Regolamento (CEE) n. 4256/88 del Consiglio del 19 dicembre 1988 recante disposizioni d'applicazione del regolamento (CEE) n. 2052/88 per quanto riguarda il FEAOG, sezione orientamento. In *Gazzetta ufficiale delle Comunità europee*. L 374, 31 dicembre 1988 (31° anno).

Grazie a questa riforma i fondi strutturali vengono integrati in una politica di coesione generale e le risorse a loro destinate aumentano, passando dal 16% al 31% circa del bilancio comunitario (64 miliardi di ECU per il periodo di spesa 1989-1993).¹⁴

Vengono introdotti anche alcuni principi fondamentali per l'utilizzo dei fondi:

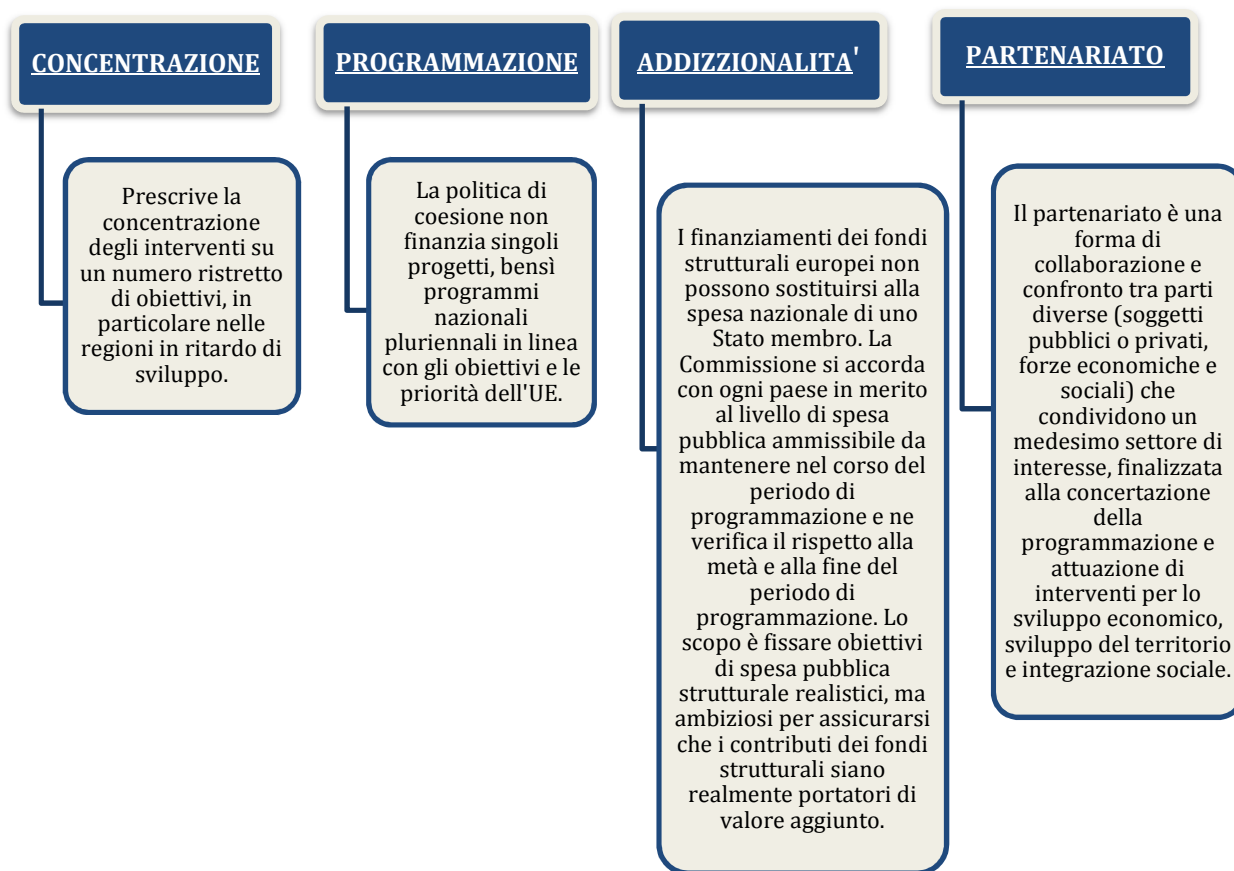


Grafico 1: Principi fondamentali per l'utilizzo dei fondi strutturali. FONTE: Ricostruzione grafica pers.

¹⁴ ŠPIDLA, V., 2008. 1989-93: Dai progetti ai programmi. In *Inforegio Panorama*. La politica di coesione dell'UE 1988-2008: investire nel futuro dell'Europa, n. 26, (giugno 2008), Pag. 9. <http://ec.europa.eu/regional_policy/sources/docgener/panorama/pdf/mag26/mag26_it.pdf>, (Ultimo accesso: 15-06-2018).

Sempre nel 1988 la Riforma definisce cinque obiettivi prioritari che la politica strutturale comunitaria intende perseguire:

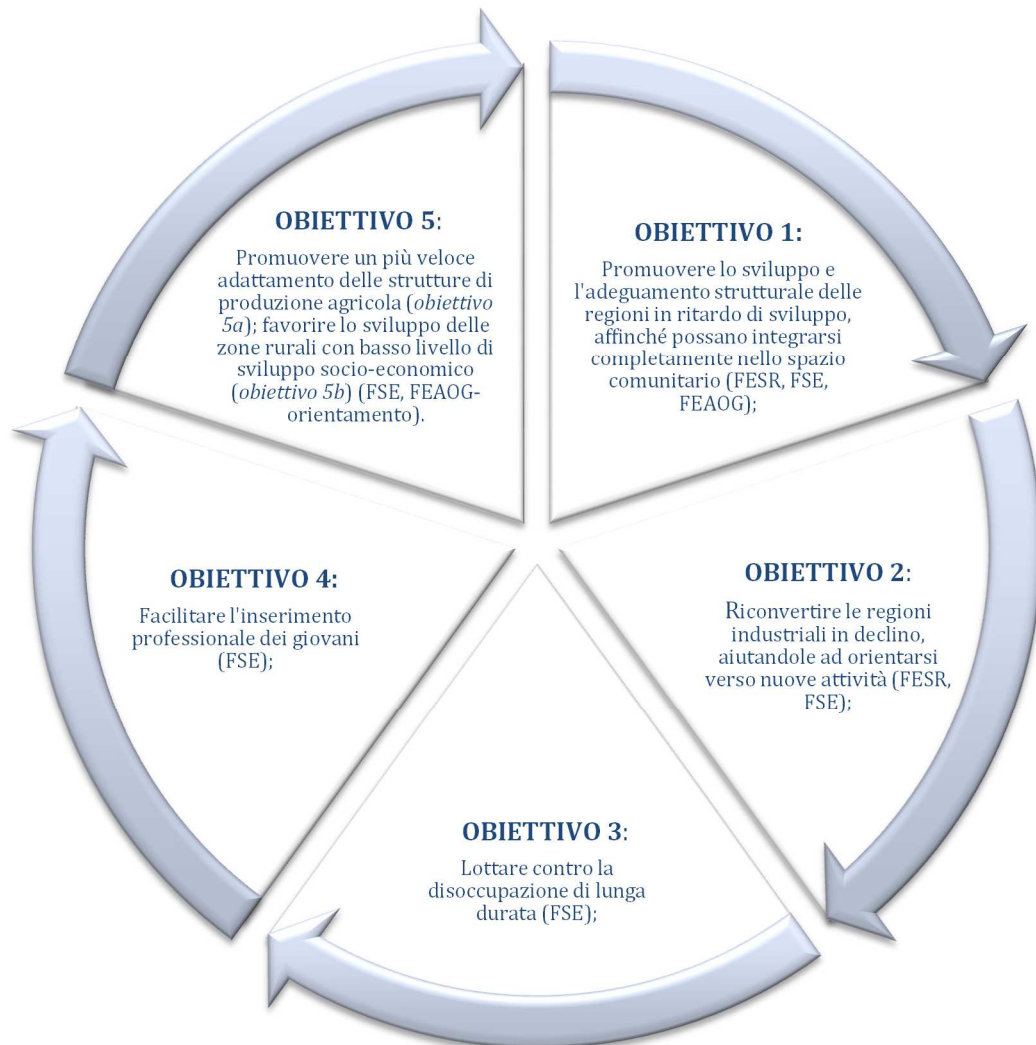


Grafico 2: Riepilogo dei Cinque Obiettivi Prioritari attuati dalla Politica Strutturale Comunitaria alla fine degli anni'80. Fonte: Ricostruzione grafica personale.

Il 7 Febbraio 1992 rappresenta un nuovo periodo di svolta grazie al *Trattato sull'Unione Europea (TUE)* firmato a *Maastricht*, insieme alla versione modificata del *Trattato* che istituisce la *Comunità Europea (TCE)*, ed entrato in vigore il 1° novembre 1993.

Con questo Trattato viene prevista la creazione di un nuovo strumento finanziario, il *Fondo di Coesione*,¹⁵ e di due nuovi organi consultivi di Commissione e Consiglio europeo, il

Comitato Economico e Sociale e il *Comitato delle Regioni*, nonché l'introduzione del Principio di Sussidiarietà.

- Il *Comitato Economico e Sociale* è composto dai rappresentanti delle forze economiche e sociali: produttori, operai, commercianti e artigiani, liberi professionisti, agricoltori, ecc.;
- Il *Comitato delle Regioni* è composto da rappresentanti delle collettività regionali e locali.

Il *Consiglio Europeo*, nel dicembre del 1992, destina ai fondi strutturali e al *Fondo di Coesione*, 168 miliardi di ECU,¹⁶ per il periodo 1994-1999, raddoppiando le risorse annue e arrivando a coprire un terzo dell'intero bilancio comunitario. In seguito all'adesione di *Austria*, *Finlandia* e *Svezia*, nel 1995, entra in vigore un nuovo regolamento¹⁷ che definisce un sesto obiettivo della politica regionale Europea a favore delle regioni a bassissima densità di popolazione di *Svezia* e *Finlandia*.

Nel 1999 vengono accorpati i precedenti Obiettivi 2 e 5 nonché gli Obiettivi 3 e 4, riducendo così da sei a tre il numero degli obiettivi dei Fondi Strutturali.¹⁸ (Grafico 3).

¹⁵ Regolamento (CE) del Consiglio n. 1164/94, del 16 maggio 1994, che istituisce un Fondo di Coesione. In *Gazzetta ufficiale delle Comunità europee*. L 130, (16 maggio 1994).

¹⁶ ŠPIDLA, V., 2008. 1994-99: consolidare e raddoppiare gli sforzi. In *Inforegio Panorama*. La politica di coesione dell'UE 1988-2008: investire nel futuro dell'Europa, n. 26, (giugno 2008), Pag. 17. <http://ec.europa.eu/regional_policy/sources/docgener/panorama/pdf/mag26/mag26_it.pdf>, (Ultimo accesso: 15-06-2018).

¹⁷ Regolamento (CEE) n. 2081/1993 del Consiglio del 20 luglio 1993 che modifica il regolamento (CEE) n. 2052/88 relativo alle missioni dei Fondi a finalità strutturali, alla loro efficacia e al coordinamento dei loro interventi e di quelli della Banca europea per gli investimenti e degli altri strumenti finanziari esistenti. In *Gazzetta ufficiale delle Comunità europee*. L 193, (31 luglio 1993).

¹⁸ Regolamento (CE) n. 1260/1999 del Consiglio del 21 giugno 1999, Regolamento del Consiglio recante disposizioni generali sui Fondi strutturali.

I tre nuovi obiettivi sono:

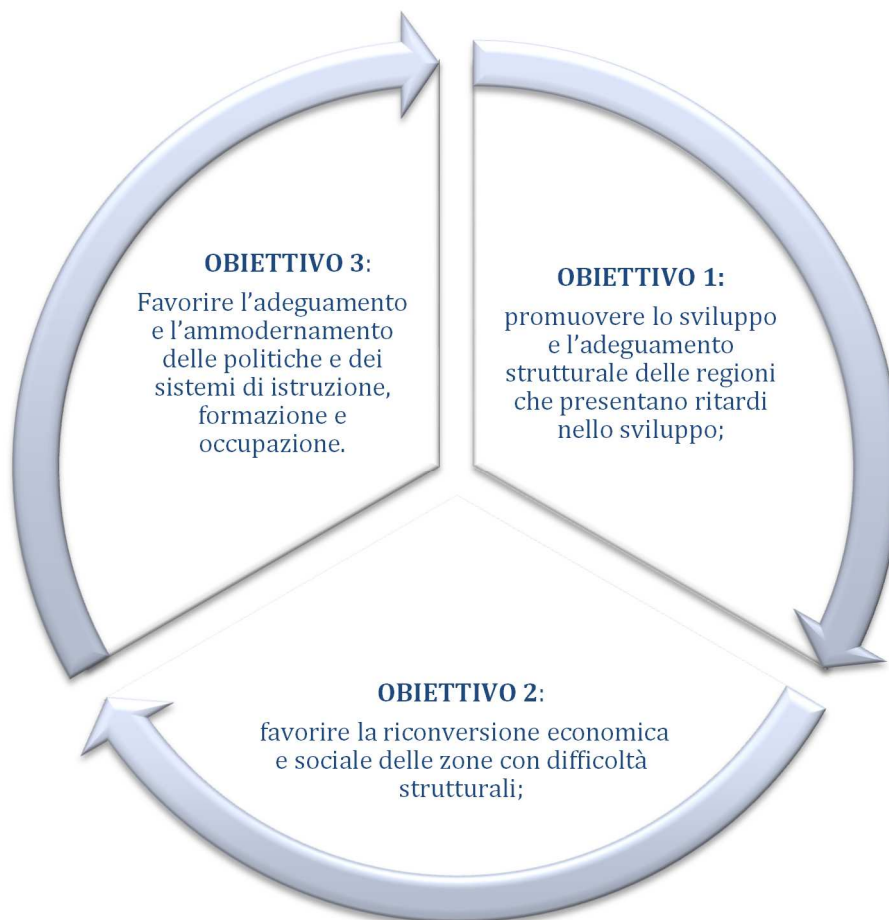


Grafico 3: In riferimento al Grafico precedente, riepilogo della fusione degli obiettivi 2 e 5, nell'obiettivo 1; mentre gli obiettivi 3 e 4 si fondono nel numero 2. Rimane invariato l'obiettivo numero 3. Fonte: Ricostruzione grafica pers.

Nel 2004 sono dieci i nuovi Stati membri aderiscono all'Unione, con un aumento della popolazione del 20% e del PIL del 5%, accrescendo le differenze già esistenti tra le regioni europee.

Nel 2006 l'approvazione di una serie di regolamenti da parte di Consiglio e Parlamento riforma nuovamente la Politica Regionale Europea.

Viene prevista la fusione degli Obiettivi 2 e 3, con la definizione di un nuovo terzo obiettivo¹⁹

Nel periodo di programmazione 2007-2013, gli obiettivi prioritari diventano:

➤ **CONVERGENZA (FESR):** ne fanno parte 84 regioni in 17 Stati membri, circa 170 milioni di persone. Si aggiungono, inoltre, altre 16 regioni (16,4 milioni di abitanti) in regime di sostegno transitorio (*phasing-out*), ovvero quelle con un PIL di poco superiore al 75%, ma per l'effetto statistico causato dall'allargamento a 27 membri.

✓ Le risorse destinate a quest'obiettivo arrivano a 282,8 miliardi di euro (programmazione 2007-2013).²⁰

➤ **COMPETITIVITÀ REGIONALE E OCCUPAZIONE (FSE):** ne fanno parte 168 regioni, circa 314 milioni di persone. Tra le 168 regioni tredici (19 milioni di persone), le cosiddette regioni a sostegno transitorio progressivo (“*phasing-in*”), ricevono stanziamenti finanziari speciali in virtù della loro precedente ammissibilità all'Obiettivo 1.

✓ Sono stati destinati 54,9 miliardi di euro (programmazione 2007-2013).²¹

¹⁹ Regolamento (CE) n. 1083/2006 del Consiglio dell'11 luglio 2006, recante disposizioni generali sul Fondo europeo di sviluppo regionale, sul Fondo sociale europeo e sul Fondo di coesione e che abroga il regolamento (CE) n. 1260/1999. In *Gazzetta ufficiale delle Comunità europee*. L 210, (25° anno), 31 luglio 2006.

²⁰ ŠPIDLA, Vladimír, 2008. 2007-13: crescita e occupazione al centro degli interventi. In *Inforegion Panorama*, n. 26, (giugno 2008), Pag. 23.

²¹ *Ivi*, pag. 24

➤ **COOPERAZIONE TERRITORIALE EUROPEA (FESR)**: prevede 13 programmi di cooperazione transnazionale. Tutti i cittadini dell'UE rientrano in almeno uno di questi.

✓ L'importo finanziario destinato a quest'obiettivo ammonta a 8,7 miliardi di euro²² (programmazione 2007-2013).

Tali obiettivi mettono a disposizione 347 miliardi di Euro per tutte le regioni europee, ripartiti tra 82% all'obiettivo **Convergenza**, 16% all'obiettivo **Competitività Regionale e occupazione** e 2% all'obiettivo **Cooperazione territoriale Europea**, (Grafico 4).

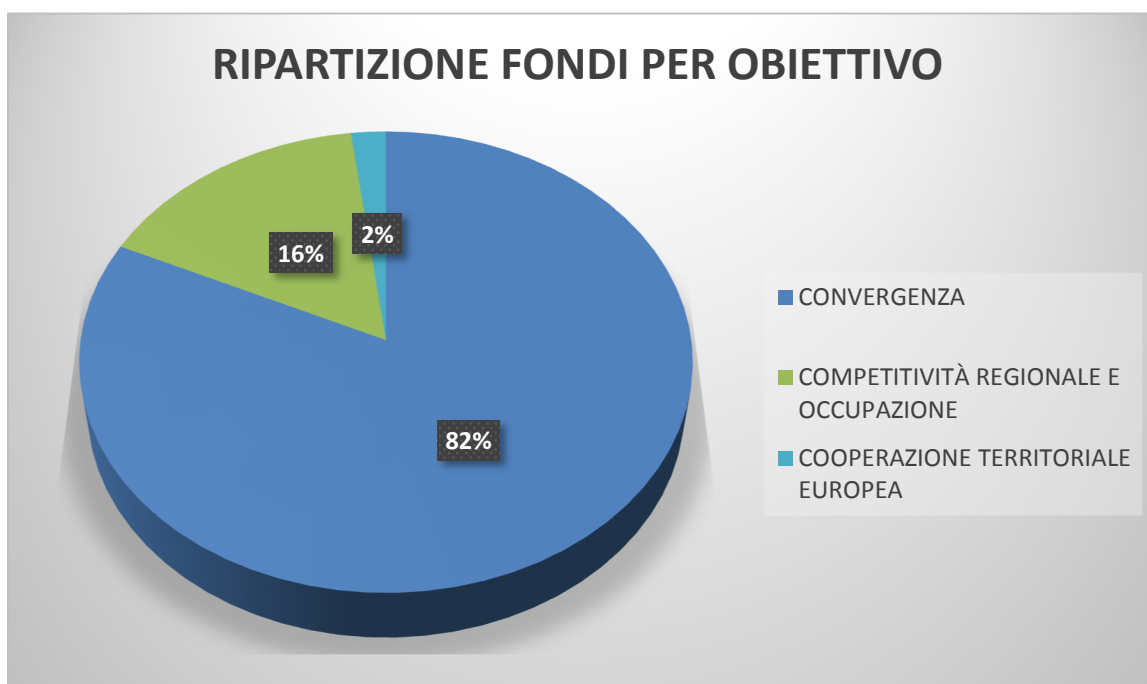


Grafico 4: Rappresentazione grafica della ripartizione dei 347 miliardi di Euro ripartiti per ogni obiettivo. Fonte: Ricostruzione grafica personale.

²² *Ibidem*

2.3 Fondi Strutturali Europei

Per ridistribuire e riequilibrare all'interno del territorio Europeo le risorse sono stati creati degli strumenti finanziari gestiti dalla Commissione Europea, ovvero i Fondi Strutturali. Pur rientrando nel bilancio dell'Unione Europea, tali fondi implicano delle modalità di spesa che prevedono una ripartizione delle competenze sia da parte della *Commissione Europea* e sia dagli Stati membri.

La Politica Regionale Europea è finanziata prevalentemente attraverso tre fondi:

- **Fondo Europeo di Sviluppo Regionale (FESR)**: creato nel 1975, promuove l'adeguamento strutturale delle regioni in ritardo di sviluppo e la riconversione delle zone industriali in declino. In tale ottica, concorre alla promozione di un alto livello occupazionale, delle pari opportunità e della protezione e dello sviluppo sostenibile. Contribuisce, inoltre, al finanziamento di interventi destinati a rafforzare la coesione economica e sociale eliminando le principali disparità regionali. In tal modo, il **FESR** dà attuazione alle priorità comunitarie e in particolare all'esigenza di rafforzare la competitività e l'innovazione, creare e mantenere posti di lavoro stabili e assicurare lo sviluppo sostenibile. Il campo d'applicazione dell'intervento del FESR rientra nel contesto degli obiettivi "Convergenza" e "Cooperazione territoriale europea".

- **Fondo Sociale Europeo (FSE)** è il principale strumento utilizzato dall'UE per sostenere l'occupazione, aiutare i cittadini a trovare posti di lavoro migliori e assicurare opportunità lavorative più eque per tutti. A questo fine, l'**FSE** investe nel

capitale umano dell'Europa: i lavoratori, i giovani e chi è alla ricerca di un lavoro. Stati membri, Parlamento europeo e Commissione europea decidono il budget dell'**FSE** e quale strategia adottare. Vengono elaborati programmi operativi di durata settennale da ogni Stato membro con la collaborazione della Commissione che verranno poi messi in pratica tramite diversi tipi di organizzazioni attive nel settore pubblico o privato.

- **Fondo di Coesione:** sono aiuti dell'Unione Europea con la finalità di finanziare infrastrutture di trasporto, progetti di educazione, salute ed ecosistema, destinati agli Stati membri con un *Reddito Nazionale Lordo (RNL)* pro capite inferiore al 90% della media dell'Unione europea. I suoi obiettivi sono la riduzione delle disparità economiche e sociali e la promozione dello sviluppo sostenibile.

Nel 2002, inoltre, viene creato il *Fondo di Solidarietà dell'Unione Europea (FSUE)*, il quale interviene in caso di calamità naturali ed esprimere la solidarietà europea alle regioni colpite all'interno dell'UE. Il Fondo è stato istituito a seguito delle gravi inondazioni che hanno devastato l'Europa centrale nell'estate del 2002.

Da allora è stato utilizzato ben 63 volte in risposta a diversi tipi di catastrofi, tra cui inondazioni, incendi forestali, terremoti, tempeste e siccità. Finora sono stati erogati oltre 3.7 miliardi di euro a favore di 24 paesi europei.

Inoltre, nel 2007 sono stati istituiti quattro nuovi strumenti finanziari:

- **JASPERS** (*Joint Assistance to Support Projects in European Regions* - Assistenza congiunta a sostegno dei progetti nelle regioni europee) è uno strumento di assistenza tecnica, è rivolto agli ultimi dodici paesi entrati nell'Unione Europea e fornisce

loro il supporto necessario a elaborare progetti che verranno poi co-finanziati dai fondi UE.

- **JASMINE** (*Joint Action to Support Micro-finance Institutions in Europe* - Azione comune a sostegno degli istituti di microfinanza in Europa) fornisce sia assistenza tecnica, sia sostegno finanziario a erogatori di microcredito non bancari e li aiuta a migliorare la qualità delle loro attività. JASMINE cerca anche di promuovere le buone prassi nel campo del microcredito e di redigere un codice di buona condotta. L'UE intende per "microcredito" un prestito inferiore ai 25000 euro.

- **JEREMIE** (*Joint European Resources for Micro to Medium Enterprises* - Risorse europee comuni per le micro, le piccole e le medie imprese) è un'iniziativa sviluppata in collaborazione tra la Commissione europea e il Fondo europeo per gli investimenti, creato con il fine di aiutare le micro, le *Piccole E Medie Imprese (PMI)* ad accedere al credito mediante i fondi strutturali.

- **JESSICA** (*Joint European Support for Sustainable Investment in City Areas* - Sostegno europeo congiunto per investimenti sostenibili nelle aree urbane) è un'iniziativa della Commissione europea realizzata in collaborazione con la Banca europea per gli investimenti (BEI) e la Banca di sviluppo del Consiglio d'Europa (CEB). L'obiettivo è promuovere investimenti sostenibili nelle aree urbane europee e portare crescita e occupazione investendo nelle infrastrutture urbane e i progetti di edilizia popolare, aiutando la riqualificazione delle aree dismesse e l'efficienza energetica.²³

²³ EUROPEAN COMMISSION. EU_REGIONAL POLICY INFOREGIO. *Financial Engineering Background: Jaspers-Jeremie-Jessica-Jasmine*, <http://ec.europa.eu/regional_policy/archive/funds/2007/jjj/index_en.htm>, (Ultimo accesso 23-06-2018).

2.4 L'evoluzione del Bilancio nel corso degli anni

Nell'arco degli anni i fondi strutturali non hanno rivestito sempre la stessa importanza. Analizzando annualmente il bilancio Europeo, appare subito molto chiaro come nei primi anni di vita delle istituzioni comunitarie solo una parte minima delle risorse europee veniva destinata ai fondi strutturali e alla politica regionale.

Il seguente grafico (5), mostra l'evoluzione di fondi strutturali rispetto al bilancio totale dell'UE:

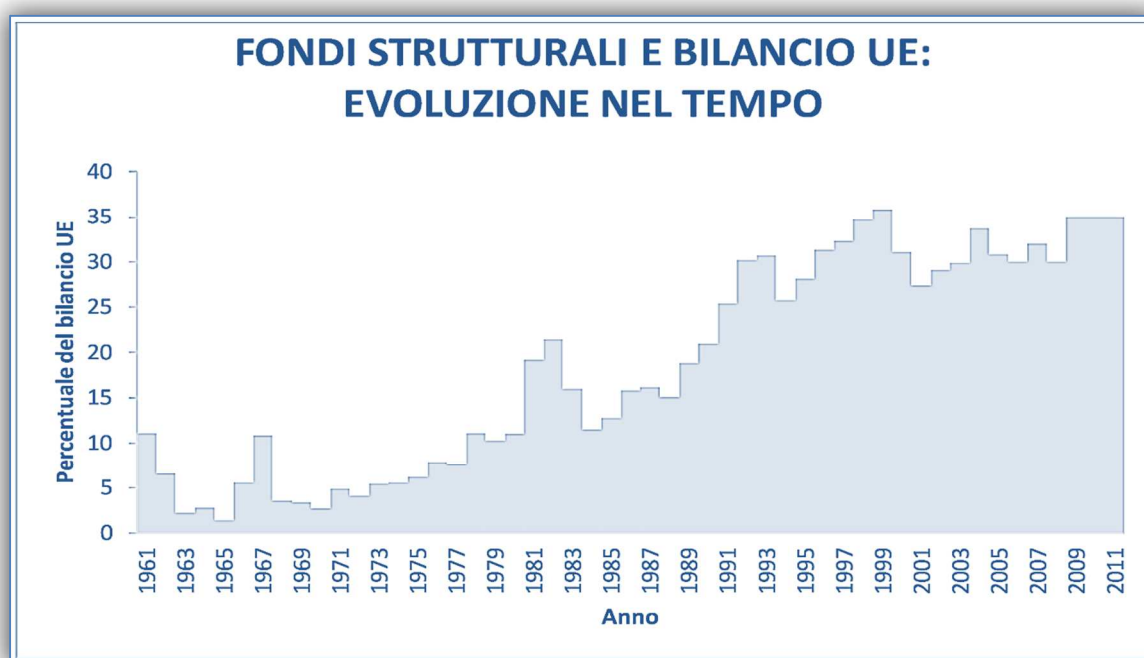


Grafico 5: Rappresentazione dell'evoluzione dei fondi strutturali rispetto al bilancio dell'Unione Europea dal 1961 al 2011. Fonte: Rappresentazione grafica pers.

Dal 1961 al 1975 la percentuale del bilancio indirizzata a questo scopo era in media del 5%. Da questo momento inizia, però, un deciso trend di crescita annuale che porta questa percentuale dal 2,7 % del 1970 fino al raggiungimento del picco massimo del 21,5 % nel 1982, portando addirittura a quadruplicare il dato, del 1972, al 4,1 %.

Questo andamento assolutamente positivo si inverte bruscamente nel 1983, quando la percentuale di risorse indirizzate ai fondi strutturali scende al 16 % del bilancio totale e ancora più nel 1984, anno in cui con 11,5 % viene quasi dimezzato il dato di due anni prima. Fortunatamente, dopo questo calo, la crescita diventa costante. Grazie alla riforma dei fondi strutturali e alla politica regionale europea del 1988 e alla successiva firma del trattato di Maastricht nel 1992, la politica per la coesione conquista un ruolo di primo piano.

A tal proposito è significativo come nel periodo di programmazione 1989-2003 la percentuale del bilancio rivolta ai fondi strutturali passa dal 16% circa del periodo precedente al 31%.

Un nuovo picco massimo viene raggiunto nel 1999, quando più di un terzo del bilancio UE riguarda la politica regionale, ovvero il 35,8%. Negli anni successivi, escludendo il 27,4% del 2001, la proporzione resta in media sempre la stessa, confermando l'importanza della politica regionale per la coesione nell'ambito degli scopi perseguiti dalle istituzioni dell'Unione Europea.

2.5 L'impatto dei Fondi di Coesione in Italia

Negli ultimi anni le disparità di reddito pro capite tra i paesi dell'Unione Europea si sono ridotte, ma a tali dinamiche si contrappone il mantenimento di significativi divari tra le regioni. Sebbene marcate differenze siano presenti in molti paesi europei (*Germania, Francia, Regno Unito, Grecia, Spagna, Belgio*) nei livelli di sviluppo regionale, l'*Italia* rappresenta un caso emblematico a causa della coesistenza di un'area del paese, il Centro Nord, molto ricca e un'area, il *Mezzogiorno d'Italia*, in cui i ritardi di crescita sono rilevanti e si perpetuano nel tempo.

Una delle cause che può rendere inefficace la relazione tra l'economia e la qualità istituzionale può essere ricercata nella storia. L'unificazione del territorio italiano è avvenuta solo nel 1861, dopo periodi di lunghe dominazioni diverse in tutta la penisola, che possono aver influito sul differente sviluppo nelle varie regioni odierne.

La storia italiana è stata oggetto di politiche ed economie talmente differenti nelle varie parti del paese, da dar vita a culture eterogenee. Ciò non giustifica il fatto che, oggi, nonostante la presenza di ingenti risorse finanziarie dedicate alla politica di coesione, il divario tra le varie regioni italiane è ancora molto marcato.

I fattori che influenzano maggiormente tale divario riguardano soprattutto la gestione della spesa, i meccanismi di allocazione delle risorse e le strategie di sviluppo delle politiche.

A tal proposito ci sorge spontaneo domandarci l'efficacia di questi fondi nel territorio Italiano. Numerosi studi sono stati svolti per appurare l'efficacia o meno dei fondi strutturali europei, ma il dibattito resta ancora molto aperto.

Dall'analisi dei dati empirici, infatti, emergono opinioni diverse e molto spesso contrastanti:

- da una parte ci sono coloro che pensano che le risorse europee abbiano contribuito alla crescita delle regioni in ritardo e siano state determinanti per far diminuire o almeno non aumentare le

differenze (Commissione Europea; Crespo-Cuaresma;²⁴ Beugelsdijk e Eijffinger²⁵);

- dall'altra, invece, gli studi che hanno evidenziato una completa inefficacia della politica perseguita dalle istituzioni europee (Boldrin e Canova²⁶, Aiello e Pupo²⁷) oppure hanno sostenuto che l'efficacia è condizionata da determinati fattori (Ederveen, de Groot, Nahuis).

Purtroppo, raramente tali studi sono letti con attenzione critica, figuriamoci se utilizzati per sostenere decisioni politico-programmatiche. La verità di fondo è che l'interesse per la valutazione nel nostro paese resta purtroppo ancora limitato.

Una cosa certa è che l'*Italia*, oggi, risulta il secondo Stato in UE per Fondi Strutturali ricevuti da *Bruxelles*, ma è sestultima su 28 per utilizzo dei soldi ricevuti.

Ripercorrendo i vari periodi di programmazione partendo dal 2000/2006, i Fondi Europei destinati al territorio italiano sono stati attribuiti alle zone che erano in ritardo a livello di sviluppo, quali: *Campania, Calabria, Sicilia, Puglia, Sardegna, Basilicata e Molise*.

I risultati sono stati molteplici tra cui: il sostegno di 2.000 imprese nelle regioni italiane, la realizzazione di nuovi impianti di smaltimento dei rifiuti, la modernizzazione e la costruzione di strade e la valorizzazione del patrimonio culturale nel Sud Italia.

A supporto del programma, l'*Italia* ha ricevuto 29 miliardi di euro per l'aiuto (**FERS** e **FSE**), somma che rientra negli obiettivi di Convergenza, Competitività Regionale e Occupazione, e Cooperazione Territoriale Europea. Le regioni che sono oggetto di Convergenza territoriale sono *Campania, Puglia, Sicilia e Basilicata*, con a

²⁴ CRESPO CUARESMA, Jesus, RITZBERGER-GRÜNWALD, Doris and SILGONER Maria Antoinette, 2008. Growth, Convergence and EU Membership. In *Journal Applied Economics* Volume 40, Issue 5, Pp. 643-656.

²⁵ BEUGELSDIJK, Maaïke, and EIJFFINGER, C. W., Sylvester, 2005. The Effectiveness of Structural Policy in the European Union: An Empirical Analysis for the EU-15 in 1995-2001. In *Journal of Common Market Studies*, Vol. 43, No. 1, Pp. 37-51, (March 2005).

²⁶ BOLDRIN, Michele, CANOVA, Fabio, 2001. Inequality and Convergence in Europe's Regions: Reconsidering European Regional Policies. In *Economic Policy* N. 32, Pp. 53-75.

²⁷ AIELLO, Francesco, PUPO, Valeria, 2009. The Impact of European Cohesion Policy. A Study on Italian Regions. In *Rivista Italiana degli Economisti*, Vol. 3, N. 12, 2009. Pp. 421-454.

disposizione più del 75% dei Fondi Italiani. Il resto d'Italia è coinvolto nell'obiettivo della Competitività e Occupazione.



Fig. 2: Suddivisione degli obiettivi di Convergenza e Competitività per l'impiego dei fondi 2007-2013, in Italia. Fonte: Programma Operativo Nazionale, pag. 1.

Si definisce, così, il *Quadro Strategico Nazionale* che stabilisce i principi attuativi della Politica di Coesione attraverso una serie di priorità regionali per l'impiego dei Fondi 2007/2013.

La spesa certificata alla *Commissione Europea* indica lo stato di avanzamento finanziario impegnato nella Politica di Coesione e di Sviluppo, diventando così una variabile fondamentale per il disimpegno automatico. L'avanzamento finanziario è calcolato come una quota percentuale rispetto alla spesa certificata ed è in relazione ad una certa data, in rapporto alla dotazione finanziaria disponibile totale.

Per quanto riguarda l'analisi che è stata condotta a fine dicembre 2012, attraverso i dati del Dipartimento per lo Sviluppo e la Coesione Economica, si evidenzia un'incapacità italiana di gestione dei fondi **FESR** e **FSE**.

Come si può facilmente intuire, il de-finanziamento delle risorse comunitarie si traduce in una riduzione anche del finanziamento nazionale. Quindi i Fondi devono essere impegnati in un programma entro la fine del 2013 e necessariamente spesi alla fine del 2015.

Come precedentemente detto, l'Italia si pone in uno degli ultimi posti, assieme alla *Bulgaria* e alla *Romania*, nella classifica dei paesi in via di sviluppo.

La difficoltà però che il nostro Paese riscontra nell'investimento della spesa prevista dai programmi europei è data da un ritardo e da una lentezza storica nei primi anni della programmazione, che sembra quasi essere dovuta ad un fattore sistematico, che va a ripetersi negli anni. Si può notare come quasi un terzo della spesa della programmazione sia stato concentrato nelle ultime fasi della programmazione stessa.

Il ritardo cronico dell'amministrazione rende insostenibile il ritmo di spesa e di sviluppo in quanto limita l'impegno quasi esclusivamente nell'ultimo periodo del settennato. Tale situazione mette in risalto la criticità della *Governance* italiana rispetto ai Paesi Europei, minando anche la priorità dell'azione per migliorare l'impatto dei Fondi.

Si rimarca che la situazione italiana genera una scarsa capacità di gestione dei Fondi destinati alle aree Convergenza e Competitività. Inutile dire che l'esperienza acquisita in vent'anni di programmazione alla Coesione e Sviluppo, si presenta inefficiente e inefficace, collocandosi addirittura dopo la *Romania* (per le regioni di

obiettivo Convergenza, spendendo il 16%) e dopo la *Repubblica Ceca* (per le regioni di obiettivo Competitività, spendendo il 26%), che sono di recente entrata nell'UE.

Risultati relativamente più incoraggianti vengono registrati, invece, nella capacità di spesa del **FSE**.

L'*Italia*, per quanto riguarda l'area Convergenza, ha sfruttato solamente il 20% di spesa rispetto alla dotazione totale, collocandosi come una delle ultime regioni assieme a *Malta*, *Bulgaria*, *Romania* e *Grecia*; in contrapposizione la *Lettonia*, l'*Austria* e il *Portogallo* hanno dimostrato un'ottima capacità di gestione della spesa.

Purtroppo anche nell'Obiettivo di Competitività Regionale e Occupazionale l'*Italia*, assieme alla *Francia*, alla *Danimarca*, all'*Olanda* e al *Regno Unito*, occupa uno degli ultimi posti della classifica Europea, presentando la stessa dinamica di spesa con cui condivide appunto gli ultimi posti e spendendo solo 30% dei Fondi a disposizione.

3.1 Definizione e cenni storici della *Content Analysis*

La *Content Analysis* nasce nel campo delle scienze sociali ed è definita come “*A method of analysis of communication message using as a systematic system of counting items within established categories*”²⁸.

Si tratta di un metodo di ricerca che, partendo da informazioni qualitative (i testi), ci permette di studiare, sistematicamente e oggettivamente, i contenuti di un processo di comunicazione, applicando sistemi di classificazione del contenuto con l’obiettivo di pervenire a dati quantificabili da sottoporre ad analisi statistiche. La *Content Analysis* può essere applicata in diversi campi, quali: testi orali o scritti (discorsi, documenti, giornali, libri) o contenuti comunicativi di altro genere (cinema, musica, tv, prodotti radiofonici).

La ricostruzione delle radici storico-concettuali di questo metodo di ricerca è di fondamentale importanza per individuare i limiti ma anche le potenzialità che caratterizzano il più tradizionale degli approcci metodologici all’analisi dei testi. Tale studio prende il via verso la fine del Settecento. Al tempo, la chiesa temeva che la diffusione della stampa potesse portare alla diffusione di temi non religiosi.

Alla fine del XVIII secolo lo studioso **Klaus Krippendorff**,²⁹ individua in *Svezia* il primo caso di documento di analisi quantitativa di materiale di stampa: si trattava di una raccolta di inni religiosi chiamati i “Canti di Sion”.

Tale analisi era stata ordinata dalla Chiesa in quanto sospettava la presenza di contenuti eterodossi in quei testi.

A partire da questa prima applicazione pionieristica di tale metodo, bisogna aspettare le due guerre mondiali per documentarne un impiego su larga scala.

²⁸ HOLSTI, Ole, 1968. Content Analysis. In LINDZEY, G. and ARONSON, E. (Eds), The Handbook of Social Psychology, Vol 2. Hoboken, New Jersey, John Wiley & Sons Inc.

²⁹ KRIPPENDORFF, Klaus. 2004. Reliability in content analysis: some common misconceptions and recommendations. In Human Communication Research, Vol. 30(3), Pp. 411-433.

L'insieme di tecniche che prende il nome di *Content Analysis* nasce, effettivamente, negli Stati Uniti, intorno agli anni Venti, nell'ambito della sociologia delle comunicazioni di massa e degli studi sulla propaganda politica, in particolare, sull'aspetto persuasivo e politico-ideologico dei messaggi veicolati dalla comunicazione politica del tempo, sia sul versante nazionale che internazionale.

La *Content Analysis* affonda le proprie radici storiche e teoriche nel noto paradigma comunicativo elaborato da **Harold Lasswell**, che ha fornito uno schema descrittivo dell'atto comunicativo per lungo tempo dominante nella *Communication Research*, utilizzato per l'analisi della trasmissione e della ricezione dei messaggi.³⁰

Seguendo tale approccio ogni possibile tentativo di analisi e di ricerca sui media si risolve e si riduce nella risposta a cinque interrogativi-chiave:

1. **Chi dice** (*Who says*);
2. **Che cosa** (*What*);
3. **Attraverso quale canale** (*through Which channel*);
4. **A chi** (*to Whom*);
5. **Con quale effetto** (*with Which effects*).

A ciascun quesito può essere associato lo sviluppo di uno specifico filone di ricerca: il primo definisce l'analisi degli emittenti, il secondo l'analisi del contenuto dei messaggi (*Content Analysis*), il terzo l'analisi dei mezzi tecnici, il quarto l'analisi dell'audience e l'ultimo lo studio degli effetti della comunicazione (Cfr. Morcellini-Fatelli).³¹

Nei primi anni Cinquanta, la *Content Analysis* si è arricchita di ulteriori contributi, sia teorici che metodologici, di altre discipline come la psicologia; nel corso

³⁰ LASSWELL, Harold Dwight, 1949. *The Language of Power*, In LASSWELL, H.D. e LEITES, Nathan, (Eds.), *Language of Politics; Studies in Quantitative Semantics*. New York, G.W. Stewart Publisher Inc. Pp-3-19.

³¹ MORCELLINI, Mario e FATELLI, Giovanbattista, 1999. *Le Scienze della Comunicazione. Modelli e percorsi disciplinari*. Roma, Carocci Editore.

del Novecento ha esteso il proprio campo di applicazione a forme di comunicazione non scritte, come l'iconografia e i prodotti audiovisivi.

Con lo sviluppo di programmi informatici poi, nei primi anni Sessanta, l'analisi del contenuto si è dotata di supporti tecnici sempre più raffinati per l'elaborazione dei dati.

Il **web** si è rivelato uno strumento cruciale, che ha incrementato le potenzialità della *Content Analysis*, rendendo accessibili archivi online e database di quotidiani, tv e radio di tutto il mondo.

3.2 Fasi della *Content Analysis*

L'analisi del contenuto segue diverse fasi. Secondo **Wimmer** e **Dominick** (1997) sono nove le fasi importanti da applicare in questo processo.³²

- 1) **Formulare l'ipotesi o la domanda di ricerca.** Come per ogni altra forma di ricerca, occorre avere ben chiara un'ipotesi o domanda di ricerca, maturata normalmente dopo un attento esame della letteratura già esistente sull'argomento. Un'ipotesi di ricerca ben formulata facilita lo sviluppo di categorie di contenuto accurate e significative che a loro volta aiutano a produrre dati di valore.
- 2) **Definire la popolazione considerata.** Il ricercatore deve stabilire i confini del contenuto preso in esame e ciò richiede un'adeguata definizione della porzione di popolazione interessata. Prima di decidere tale popolazione, occorre

³² WIMMER, Roger and DOMINICK, Joseph, 1997. *Mass media research. An introduction*. Belmont (CA), Wadsworth Publishing Company.

determinare il contesto considerando due dimensioni: l'area topica che sarà studiata e il periodo di tempo dello studio.

3) **Selezionare un campione.** Se la ricerca si concentra su una quantità finita di dati, il contenuto si può facilmente misurare. Tuttavia, nella maggior parte dei casi il contenuto preso in esame è troppo vasto per cui bisogna selezionare un campione. La tecnica di campionatura più comunemente usata nell'analisi del contenuto è quella a livelli multipli. Prima il ricercatore sceglie una serie di fonti (giornali, riviste, film...) e dopo seleziona certi dati (per esempio, certi numeri di un giornale o di una rivista). Con questa selezione a doppio livello, il ricercatore può arrivare ad avere una quantità di dati da analizzare più facilmente gestibile.

4) **Selezionare l'unità di analisi.** L'elemento minimo di un'analisi del contenuto, ciò che viene effettivamente misurato, si chiama unità di analisi. Mentre con il contenuto scritto, tale unità può essere una singola parola o un simbolo, un tema o un articolo o una storia, nelle analisi della televisione e del cinema le unità di analisi possono essere personaggi, azioni, scene o interi programmi. È più semplice contare certe unità di analisi di altre. Per esempio, è piuttosto facile identificare gli avvenimenti riportati dalle notizie internazionali, mentre contare le azioni "Violente" in televisione può diventare assai complicato. Occorre dunque formulare in maniera chiara e precisa delle definizioni operative dell'unità di analisi, come pure rendere evidenti e facilmente osservabili i criteri di inclusione. Il ricercatore non arriva a questi obiettivi subito, ma solo dopo un graduale processo di 'prova ed errore' e di perfezionamento.

5) **Costruire le categorie di contenuto dell'analisi.** Questa fase probabilmente è la più importante. Come Berelson nota nel suo storico volume (1952),³³ se non si sviluppano delle categorie di contenuto ben definite e adeguate all'argomento in esame, l'analisi non sarà produttiva. Bisogna dunque classificare i dati secondo certe categorie di contenuto sviluppate dallo stesso ricercatore. Queste categorie devono essere reciprocamente esclusive, esaustive e affidabili. Innanzitutto devono essere reciprocamente esclusive, e cioè un'unità di analisi può essere inserita soltanto in una categoria di contenuto. Nel caso in cui un'unità possa essere inserita in due categorie diverse, allora il ricercatore dovrà modificare il suo sistema di categorizzazione in modo da rendere le categorie reciprocamente esclusive. Secondo, le categorie devono essere esaustive, e cioè ogni unità di analisi deve essere inserita in una categoria; in molte analisi del contenuto i ricercatori spesso risolvono il problema dell'esaustività creando una categoria chiamata 'altro' o 'vari'. Terzo, le categorie devono essere affidabili. I vari codificatori devono essere ampiamente d'accordo sulla categoria di contenuto più adeguata per ciascuna unità di analisi. Questo accordo tra i diversi codificatori viene quantificato e chiamato affidabilità inter-codificatore.

6) **Stabilire un sistema quantitativo.** Anche se la quantificazione nell'analisi del contenuto può richiedere tutti e quattro i livelli di misurazione previsti dalla ricerca delle scienze sociali, di solito ne vengono impiegati solo tre: nominale, a intervalli e dei dati proporzionali. Le misurazioni nominali permettono al ricercatore di contare semplicemente la frequenza delle unità in ciascuna categoria di contenuto. Trattandosi di una misurazione superficiale, occorre aggiungere quella a intervalli per dare più

³³ BERELSON, B., 1952. *Content analysis in communication research*. New York, The Free Press.

profondità e complessità all'analisi. I ricercatori possono sviluppare delle scale che i codificatori usano per classificare certi attributi dei personaggi o situazioni. Le scale di classificazione, tuttavia, introducono nel processo di codificazione un elemento di soggettività. Nella ricerca delle comunicazioni di massa, le misurazioni proporzionali si riferiscono allo spazio e al tempo. Per esempio, quando si analizzano i fatti riportati da un giornale, si può misurare la lunghezza del relativo testo in centimetri-colonna. Nel caso delle trasmissioni radiotelevisive si può misurare la lunghezza del programma in minuti o secondi.

7) **Codificare il contenuto.** Il processo di classificazione di un'unità di analisi all'interno di una certa categoria di contenuto si chiama codificazione e codificatori sono le persone che se ne incaricano. In qualsiasi analisi del contenuto il numero dei codificatori è piuttosto basso (da tre a sei). Per far sì che l'analisi sia più affidabile, tutti i codificatori devono sottoporsi a una o più sedute di preparazione durante le quali familiarizzano con le definizioni e le unità di analisi, e si esercitano nelle procedure di codificazione. Per rendere la codificazione più semplice e affidabile, le analisi del contenuto di solito impiegano dei prestampati standardizzati simili ai questionari usati nei sondaggi. Grazie a essi i codificatori possono classificare i dati semplicemente spuntando le voci presenti nel prestampato.

8) **Analizzare i dati.** Come i sondaggi, anche l'analisi del contenuto richiede un'analisi statistica la cui tipologia dipende dagli obiettivi della ricerca. Quando un ricercatore vuole testare l'ipotesi di partenza, occorre usare statistiche inferenziali. Viene di solito usato il test chi-quadro poiché i dati dell'analisi del contenuto tendono a essere nominali. Dal momento che di solito

vengono prese in considerazione grandi quantità di dati, si ricorre spesso a software specializzati come l'SPSS (Statistical Package for Social Science).

9) **Interpretare i risultati.** È nell'interpretazione dei risultati che un ricercatore rivela tutta la sua creatività e originalità. Oltre a presentare i risultati della sua ricerca in maniera chiara e coerente, egli deve dimostrare che essi aggiungono qualcosa di nuovo e di significativo alla letteratura già esistente sull'argomento. Per cui, in questa fase del processo, il ricercatore deve ritornare al suo studio della letteratura e dimostrare come i suoi risultati provino o smentiscano le conoscenze esistenti.

3.3 Limiti e Osservazioni

Nonostante rappresenti ormai una tecnica utilizzata in molte discipline, e possa contare su diversi metodi di applicazione codificati, restano tuttavia una serie di problemi connessi alle procedure di separazione delle unità di testo.

In generale, la *Content Analysis* viene considerata come una tecnica di ricerca ibrida che unisce l'approccio quantitativo a quello qualitativo: lo studio delle occorrenze di un termine che si ripete più frequentemente di altri presuppone che lo stesso termine rappresenti un elemento importante della comunicazione (in linguistica è definita parola-chiave); i legami esistenti tra il segno linguistico e il suo significato, ad esempio, possono restituire un'analisi qualitativa della comunicazione, in grado di cogliere il suo essere pluridimensionale.

In definitiva, l'analisi di contenuto rappresenta un approccio al testo molto generale, chi svolge la ricerca e l'analisi, raccogliendo il materiale dagli archivi, deve selezionare e categorizzare i dati che, grazie all'approccio sistematico e all'uso dei computer, possono essere anche molto numerosi.

In conclusione possiamo dire che l'analisi del contenuto può essere dispendiosa in termini di tempo e denaro. Nonostante tutto questo, se si è rigorosi per tutta la durata dello studio, l'analisi del contenuto può essere uno strumento molto efficace per approfondire la nostra conoscenza del mondo dei media e dei loro messaggi.

CAPITOLO IV

TOPIC MODELING

Con la crescente quantità di dati degli ultimi anni, per lo più non strutturati, è difficile ottenere le informazioni rilevanti e desiderate.

A tal proposito, però, la tecnologia ha sviluppato alcuni potenti metodi che possono essere utilizzati per estrarre i dati e recuperare le informazioni che stiamo cercando.

Una di queste tecniche nel campo del *Text Mining* è il *Topic Modeling*. Come suggerisce il nome, è un processo che ci permette di identificare automaticamente gli argomenti presenti in un oggetto di testo e di scoprire strutture semantiche nascoste all'interno del corpus.

Dal punto di vista tecnico il *Topic Modeling* è una tecnica di approccio non supervisionata utilizzata per trovare e osservare il gruppo di parole (chiamate "argomenti") in grandi gruppi di testi.

Gli argomenti possono essere definiti come uno schema ripetuto di termini concomitanti in un corpus.

Esistono diversi algoritmi di *Topic Modeling* ma, quello che tratterò in questo capitolo è l'algoritmo *Latent Dirichlet Allocation (LDA)*, fondato su un approccio probabilistico bayesiano.³⁴

³⁴ Una rete bayesiana (**BN**, *Bayesian Network*) è un modello grafico probabilistico che rappresenta un insieme di variabili stocastiche con le loro dipendenze condizionali attraverso l'uso di un *Grafo Aciclico Diretto (DAG)*. Per esempio una rete Bayesiana potrebbe rappresentare la relazione probabilistica esistente tra i sintomi e le malattie. Dati i sintomi, la rete può essere usata per calcolare la probabilità della presenza di diverse malattie. Il nome deriva dal teorema di Bayes, che costituisce il fondamento di questo approccio. Il teorema di Bayes prende a sua volta il nome dal reverendo Thomas Bayes (1702-1761). Formalmente le reti Bayesiane sono grafi diretti aciclici i cui nodi rappresentano variabili casuali in senso Bayesiano: possono essere quantità osservabili, variabili latenti, parametri sconosciuti o ipotesi. Matematicamente, una rete bayesiana è un grafo aciclico orientato in cui: i nodi rappresentano le variabili, gli archi rappresentano le relazioni di dipendenza statistica tra le variabili e le distribuzioni locali di probabilità dei nodi figlio rispetto ai valori dei nodi padre. Una rete bayesiana rappresenta la distribuzione della probabilità congiunta di un insieme di variabili. Cfr. TRECCANI.IT_Enciclopedia della

Presenterò inoltre uno dei Software che utilizza tale algoritmo noto come Mallet che è servito per analizzare i dati della mia ricerca spiegata in maniera dettagliata nel prossimo capitolo

4.1 Algoritmo *Latent Dirichlet Allocation*

Le tecniche di *Topic Modeling*³⁵ sono una serie di metodi di analisi testuale che ricercano le strutture tematiche nascoste all'interno di un corpus di documenti e raggruppano i documenti in base al loro argomento principale.

Una delle principali metodologie di topic modeling è il modello *Latent Dirichlet Allocation* (**LDA**) ideato da David Blei, Andrew Ng e Michael Jordan e presentato in un articolo del 2003 su *Journal of Machine Learning Research*.

È un modello di analisi del linguaggio naturale che permette di comprendere il significato semantico del testo analizzando la somiglianza tra la distribuzione dei termini del documento con quella di un argomento specifico (*Topic*) o di un'entità.

In altre parole, si assume che vi sia una distribuzione di *topic* nel *corpus* di documenti e ad ogni topic è associata una sequenza di parole che ci forniscono un'idea dei temi trattati nel corpus di documenti.

Matematica (2013), Inferenza Bayesiana, (voce in), <http://www.treccani.it/enciclopedia/inferenza-statistica-bayesiana_%28Enciclopedia-della-Matematica%29/>; KOLLER D., FRIEDMAN N., 2009. *Probabilistic Graphical Models. Principles and Techniques*. Cambridge, Massachusetts, USA, The MIT Press.

³⁵ BLEI, David M, Y.NG, Andrew, and JORDAN, Michael, I., 2003. Latent dirichlet allocation. In *The Journal of machine Learning research*, N. 3, Pp. 993–1022,

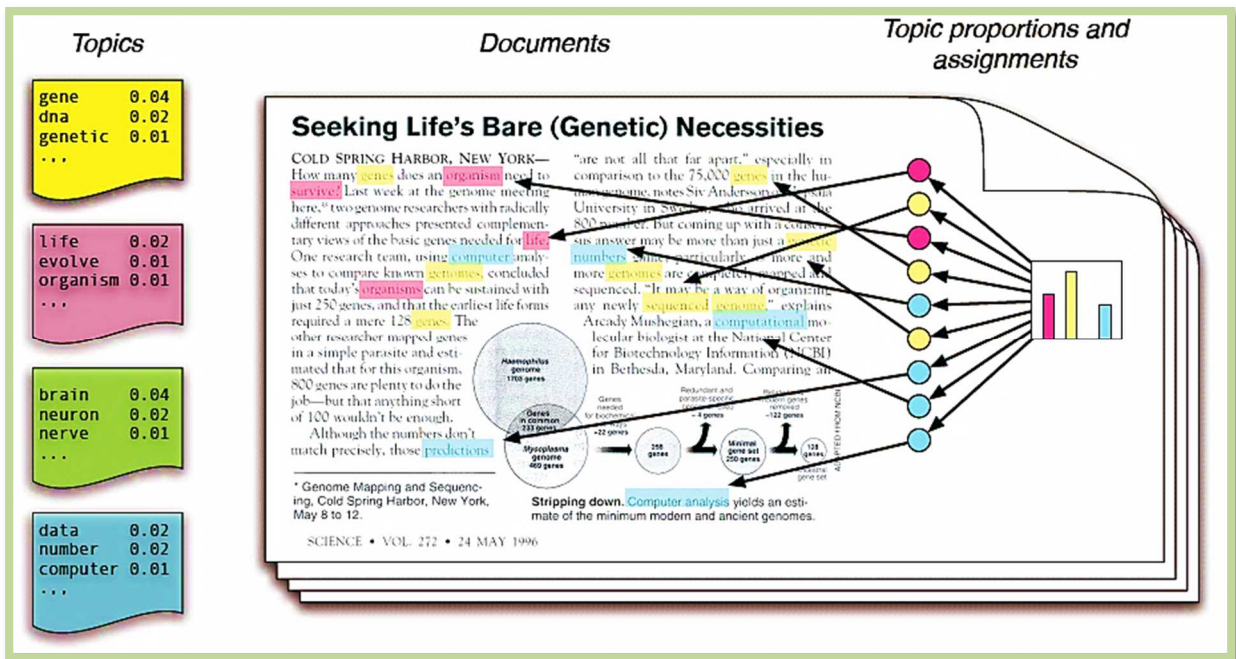


Fig. 3: Esempio di funzionamento di Algoritmo Latent Dirichlet Allocation (LDA).

Fonte: BLEI, David, M., 2012. Probabilistic topic models. In Communications of the ACM, 55(4), Pp. 77–84,

Ad esempio, prendendo un generico testo (Fig. 3) possiamo vedere che questo può essere composto da vari *topic* a cui sono associate parole specifiche di un vocabolario definito in precedenza sulla base dei testi processati e analizzati.

In particolare, nella colonna a sinistra sono riportati una serie di gruppi di parole con la probabilità più alta di appartenere ad un dato *topic*, fornendo così un'idea del tema trattato nel *topic* stesso; un documento/testo perciò risulta essere costruito mettendo insieme più *topic*.

In genere, si assume che ciascun testo sia ottenuto scegliendo prima un *topic* e solo successivamente un gruppo di parole per discutere il *topic* scelto.

A questo punto, utilizzando tecniche di statistica Bayesiana per la stima delle due distribuzioni relative ai *topic* e alle parole, si può stimare sulla base delle parole presenti in un testo, la probabilità che ciascun documento sia attribuito ad uno o più *topic*.

Esistono diversi software di *Topic Modeling*; uno degli strumenti più semplici da utilizzare per la nostra analisi è il Software **MALLET**.

4.2 *Software Mallet*

Mallet un software *open-source* sviluppato dalla *University of Massachusetts Amherst* basato sul linguaggio **Java** che utilizza un'implementazione del campionamento Gibbs, ovvero, una tecnica statistica che permette di costruire una distribuzione campione, per creare i suoi modelli tematici.

Mallet è formato da sofisticati strumenti per la **Classificazione dei Documenti**: efficienti *routine* che servono per convertire il testo in “*features*”, una serie di algoritmi (come Naïve Bayes, Maximum Entropy e Decision Trees) e con l'utilizzo di una serie di metriche contiene del codice che valuta le prestazioni del classificatore.

Oltre agli strumenti per la classificazione dei documenti ne contiene altri per il **Tagging di Sequenza**. Gli algoritmi includono i *Modelli di Markov Nascosti*, ovvero gli **HMM** (*Hidden Markov Model*), i modelli di Markov di Entropia massimi e i *Campi Casuali Condizionali a Catena Lineare*: **CRF** (*Conditional Random Field*).

Dopo aver scaricato e creato MALLET, lanciando alcuni comandi da terminale sarà possibile lavorare con il software. Per prima cosa, bisogna assicurarsi di passare alla directory MALLET prima di effettuare qualsiasi operazione.

Supponiamo, inoltre, che il formato dei nostri file da analizzare sia .txt, quindi tramite il comando sotto elencato sarà possibile trasformare tali file nel formato richiesto dal software, ovvero .Mallet.

```
./bin/mallet import-dir --input [path document] --output  
[nome_file.mallet]  
  
--keep-sequence --stoplist-file [path stopwords]
```

Tale codice contiene due comandi principali:

- **--input [path documents]:** al posto di path document bisogna inserire il percorso della cartella contenente la raccolta dei documenti in modo tale da potervi accedere.
- **--Output [nome_file.mallet]:** permette di avere come output un file .mallet contenente i documenti importati.

Il resto dei comandi sono opzionali:

- **--keep-sequence:** questa opzione conserva il documento come una sequenza di funzioni di parole, piuttosto che un vettore di conteggi di parole. Utilizzare questa opzione per le attività di etichettatura sequenziale. Il toolkit di modellazione argomento di MALLET richiede anche sequenze di caratteristiche piuttosto che vettori di caratteristiche.
- **--stoplist-file [path stopwords]:** questa opzione inoltra in mallet un file presente nel percorso da specificare in [path stopwords] contenente una lista di avverbi, congiunzioni, pronomi che mallet dovrà ignorare durante l'analisi dei documenti.

Una volta creato il nostro *file.mallet* contenente tutti i documenti, lanceremo un ulteriore comando dove andremo ad importare la raccolta precedentemente creata (*.mallet*) e ad analizzare ogni singolo file contenuto nella raccolta.

In base al numero fisso di *topic* inserito dall'utente, il software restituirà, per ognuno, una lista di parole.

Nel dettaglio, il software analizza il significato di ogni parola del file in modo tale da poter collocare ogni singola parola all'interno del *topic* più adeguato.

Questo processo dura fino alla fine dell'analisi di tutti i file in quanto le parole collocate nei *topic* possono essere modificate fino alla fine. Il comando che ci permette tutto questo è il seguente:

```
./bin/mallet train-topics --input [nome_file].mallet --num-topics [n] --num-top-words [n] --num-iterations [n] --output-state output.gz --output-topic-keys output-keys.txt --output-doc-topics output-composition.txt --diagnostics-file output.xml
```

Tale codice contiene i seguenti comandi:

- **--input[nome_file].mallet:** richiama il file precedentemente creato.
- **--num-topics[n]:** andiamo a definire il numero dei topic che vogliamo generare.
- **--num-top-words[n]:** andiamo ad indicare il numero di parole da visualizzare.
- **--num-iterations[n]:** indica il numero di iterazioni che il programma deve eseguire.

Che andranno a generare i seguenti output:

- **--output-state output.gz:** questa opzione genera un file di testo compresso contenente le parole nel corpus con le assegnazioni degli argomenti. Questo formato di file può essere facilmente analizzato e utilizzato da un software non basato su Java. Si noti che il file di stato sarà GZipped, quindi è utile fornire un nome file che termina con .gz.

- **--output-topic-keys output-keys.txt:** questo file contiene una "chiave" costituita dalle prime k parole per ogni argomento (dove k è definito dall'opzione --num-top-words). Questo output può essere utile per verificare che il modello funzioni e per visualizzare i risultati del modello. Inoltre, riporta il parametro Dirichlet di ogni argomento. Se l'ottimizzazione dell'iperparametro è attivata, questo numero sarà approssimativamente proporzionale alla porzione complessiva della raccolta assegnata a un determinato argomento.
- **--output-doc-topics output-composition.txt:** questa opzione specifica un file contenente l'elenco di tutti i nomi dei documenti della raccolta e per ognuno di essi indica con quanta incidenza il documento è composto da ciascun *topic*.

Grazie all'utilizzo di questi *software* si riescono ad analizzare mole di dati molto grandi in pochissimi minuti e ad ottenere risultati attendibili ed omogenei.



CAPITOLO V

CASO DI STUDIO

Come già osservato nelle pagine precedenti, l'obiettivo di questo lavoro è quello di sperimentare metodi e modelli per ottenere descrizioni tematiche degli elementi di un corpus di documenti.

In questo capitolo verranno utilizzate tutte le nozioni espresse in precedenza che ci permetteranno di descrivere la differenza tra due metodi di analisi eseguiti su una serie di articoli de *Il Sole 24 Ore* che coprono il periodo di tempo che va dal 18/03/2016 al 18/12/2017, aventi come argomento di studio la **Politica di Coesione Europea**.

Lo studio consiste nel descrivere i risultati ottenuti dall'analisi qualitativa di una raccolta di articoli cartacei che ci sono stati consegnati dalla *Regione Puglia* in merito a questo tema e quelli ottenuti dall'analisi condotta attraverso *Topic Modeling* di articoli scaricati tramite il sito ufficiale de *Il Sole 24 Ore* in formato digitale e analizzati successivamente utilizzando il *Software Mallet*.



Nei vari paragrafi verranno elencati e spiegati i vari passaggi e le varie operazioni eseguite riguardanti le due tipologie di analisi e in conclusione verranno studiati e confrontati i risultati ottenuti.

5.1 Analisi Qualitativa

L'oggetto dell'analisi sono una raccolta di 104 articoli de *Il Sole 24 Ore*, messi a disposizione dalla *Regione Puglia*, riguardanti la **Politica di Coesione** che coprono il periodo di tempo che va dal 18/03/2016 al 18/12/2017.

Quest'analisi è stata svolta manualmente esaminando ogni singolo articolo in cerca di tematiche e significati che ci aiutassero a classificare e a comprendere i vari articoli. L'indagine che ci siamo proposti di fare era cercare di capire l'impatto della Politica di Coesione in Europa e in Italia e il pensiero che si aggira intorno ad essa.

Per prima cosa, nell'esaminare un articolo, ci siamo basati sull'osservazione della frequenza con cui si ripetono determinati vocaboli. In altre parole abbiamo cercato di identificare la struttura del testo, per poi fare inferenze sulla base di queste ripetizioni linguistiche.

Una volta compreso l'argomento specifico affrontato nell'articolo abbiamo iniziato a classificare gli articoli per tema, man mano che si andava avanti con l'analisi il numero dei temi aumentava fino ad arrivare ad un totale di 10 temi, accuratamente scelti, che verranno elencati e spiegati successivamente.

Selezionato il tema viene ricercato all'interno dell'articolo l'attore, ovvero, colui che esprime un pensiero in merito alla tematica, che può essere un semplice giornalista o un soggetto intervistato. Il pensiero espresso può essere classificato come positivo, negativo o neutro e, all'interno di un articolo possono essere presenti più attori e più pensieri in merito al tema affrontato.

Riassumendo all'interno di un articolo andiamo a ricercare:

1. **Attore:** colui che esprime opinioni in merito al tema affrontato nell'articolo;
2. **Tema:** la tematica specifica affrontata nell'articolo;
3. **Posizione:** ovvero il pensiero dell'attore in merito al tema affrontato nell'articolo;
4. **Data** dell'articolo.

5.1.1 Risultati analisi qualitativa

Una volta analizzati i singoli articoli sulla base di tali parametri ho creato la seguente una tabella:

ATTORE	POSIZIONE	TEMA	DATA
Vincenzo Boccia (presidente di Confindustria)	P	5	05/08/16
Giorgio Santilli (giornalista)	NT	5	05/08/16
Stefan Pan (presidente del consiglio Confindustria)	P	5	05/08/16
Natale Mazzucca (presidente del comitato politiche di coesione territoriale)	P	3	
Gianni Rusconi (giornalista) Pietro Ferrari (presidente ass. startup turismo)	P	4	05/08/16
SVIMEZ	NE	5	17/11/16
Alessandro arona (giornalista)	P	5	10/08/16
Corina Cretu (commissaria politiche regionali)	NE	5	12/10/16
Antonello Cherchi (giornalista)	NE	2	16/10/16
Gianni Trovati (giornalista)	NT	2	16/10/16
Giuseppe Chiellino	NT	1	07/11/16
Giornalista	NE	7	19/12/16
	P	8	
	NT	6	
Mauro Meazza (giornalista)	NE	8	19/12/16
Lorenzo Codogno (giornalista)	NE	2	16/09/16
Maria Adele Cerizza (giornalista)	P	3	18/03/16
Giornalista	P	6	13/03/16
Flavia Landolfi (giornalista)	P	6	13/03/16
Maria Adele Cerizza (giornalista)	P	6	13/03/16
Maria Adele Cerizza (giornalista)	P	6	13/03/16
Maria Adele Cerizza (giornalista)	P	7	26/08/16
Francesca Milano (giornalista)	P	3	26/08/16

Enrico Wolleb (direttore Ismeri Europa)	NT	1	07/11/16
Alessandro arona (giornalista)	P	1	09/11/16
Dino Pesole (giornalista)	NE	1	05/08/16
Coface (colosso francese ass. dei crediti delle imprese)	NE	1	16/10/16
Ivan Lo Bello (Unioncamere) Claudio Tucci (giornalista)	P	8	25/10/16
Chiara Sumiraschi (economista Gruppo Clas) Chiara Bussi (giornalista)	P	3	24/10/16
chiara Bussi (giornalista)	NT	3	24/10/16
Micaela Cappellini (giornalisti)	P	7	17/10/16
Francesca Barbieri/Alessandro Rota Porta (giornalista)	P	7	17/10/16
Alessandro Sacrestano (giornalista)	P	7	09/09/16
Alessandro Sacrestano (giornalista)	P	8	09/09/16
Francesco Petrucci (giornalista)	P	4	09/09/16
Francesco Petrucci (giornalista)	P	4	09/09/16
Alessandro Sacrestano (giornalista)	P	3	30/12/16
Giornalista	P	6	09/11/16
Luca Orlando (giornalista)	P	4	14/10/16
Rosanna Aciero (giornalista)	P	3	14/11/16
Margot Zenni (co-fondatrice di dalani.it)/Roberto Liscia (presidente del consorzio netcomm)	P	2	17/10/16
Maria Adele Cerizza (giornalista)	P	3	16/09/16
Alberto Bonifazi (giornalista)	P	8	30/12/16
Maria Adele Cerizza (giornalista)	P	8	12/08/16
Alberto Bonifazi (giornalista)	P	7.8	02/09/16
Silvia Marzialetti (giornalista)	P	8	02/09/16
Silvia Marzialetti (giornalista)	P	7	02/09/16
Giornalista	P	6	25/11/16
Maria Adele Cerizza (giornalista)	P	4.7	16/12/16
Laura Dominici (giornalista)	P	9	08/11/16
Giornalista	P	9	08/11/16
Giornalista	Ne	3	23/01/17
Giornalista	Ne	10	11/01/17

Direttore	Nt	3	17/01/17
Giornalista/Dg	P	4	14/02/17
Giornalista	P	5.7	01/03/17
Giornalista/Storico Italiano	P	5.7	08/03/17
Giornalista	P	4.9	27/03/17
Amministratore delegato	P	7	07/04/17
Docente Universitario	P	7	12/04/17
D.g Agrò	P	1	14/04/17
Giornalista, economista	NT	1	14/04/17
Giornalista	P	3	21/04/17
Giornalista	Ne	1	27/04/17
Alberto Quadrio Curzio (Giornalista)/Ocse/Merkel	NE	1	30/06/17
Jeffrey Sachs (giornalista)	NE	1	30/06/17
Nicola Cavalluzzo/Valentina Martignoni(giornalista)	P	3	14/08/17
Carmine Fontina(giornalista)	P	5	29/08/17
Lello Naso(giornalista)	NE	5	29/07/17
Marzio Bartoloni (giornalista)	P	3	29/07/17
Dario Franceschini (ministro della cultura)	P	2.4	10/08/17
Isabella Buffacchi (giornalista)	P	2.3.7	25/08/17
Luca De Biase (giornalista) Enrico Giovannini (economista) Marco Bentivogli (segretario generale Fim-Cisl)	NE	2.8	30/08/17
Giornalista	P	4	21/08/17
Maria Adele Cerizza (giornalista)	P	4.7	21/08/17
Luca Orlando (giornalista)	P	2.4	22/08/17
Luca Orlando (giornalista)	P	6	22/08/17
Maria Adele Cerizza (giornalista)	P	3.7	23/08/17
Maria Adele Cerizza (giornalista)	P	3.7	23/08/17
Giuseppe Chiellino (giornalista)	P	1	05/09/17
Fabrizio Barca (economista)	NE	1	05/09/17
Giornalista	NE	1	08/09/17
Alessandro Sacrestano (giornalista)	NE	5	08/09/17
chiara Bussi (giornalista)	P	2.7	18/09/17

Maria Ludovica Agrò (direttore generale agenzia coesione territoriale)	P	2	18/09/17
Jyrki Katainen (vice presidente commissione UE)	P	7	29/09/17
Antonio Tajani (presidente Parlamento)	P	3	29/09/17
Giornalista	P	4	29/09/17
Giornalista	NE	3	09/10/17
Silvia Pasqualotto (giornalista)	NE	2.4	17/10/17
Silvia Pasqualotto (giornalista)	P	4	17/10/17
Tiziana Pikler (giornalista)	P	4	17/10/17
Carmine Fontina (giornalista)	NE	1	08/11/17
Carmine Fontina (giornalista)	P	1.7	08/11/17
Francesco Petrucci (giornalista)	P	3	03/11/17
Rosanna Zari (giornalista)	P	7	03/11/17
Giornalista	P	8	14/11/17
Francesca Barbieri (giornalista)	P	8	14/11/17
Raffaele Cantone (presidente autorità anticorruzione)	P	8	15/11/17
Giornalista	P	8	15/11/17
Pierpaolo Ceroli	P	4	18/12/17
Maria Adele Cerizza (giornalista)	P	7.4	11/12/17
Silvia Pasqualotto (giornalista)	P	4	28/11/17
Giornalista	P	4	28/11/17

Tab.2: sono riportati schematicamente la lista e l'analisi di ogni singolo articolo specificando attore, posizione, tema e anno di pubblicazione. Fonte: Ricostruzione personale

Come possiamo notare la tabella contiene gli elementi di analisi precedentemente citati. Andiamo adesso a specificare a quali sono i 10 temi ricavati da tale analisi:

- **TOPIC 1 -Politica di Coesione:** Questo tema descrive i vari obiettivi della politica di coesione, come, ad esempio, rafforzare la coesione economica, sociale e territoriale.

- **TOPIC 2-Situazione Italia:** Questo tema descrive la situazione italiana in merito all'utilizzo di questi fondi e ne valuta gli aspetti sia positivi che negativi.
- **TOPIC 3 -Piani e Riforme:** Questo tema viene utilizzato per pubblicare i piani importanti e le riforme approvati dal Governo italiano.
- **TOPIC 4-Startup e Progetti:** Questo tema raccoglie e descrive tutti i progetti e le *startup* ideate e create con l'utilizzo di questi fondi e l'impatto positivo che hanno avuto, soprattutto nell'ambito tecnologico.
- **TOPIC 5 -Situazione Mezzogiorno:** Questo tema si concentra sul rapporto tra fondi strutturali e lo sviluppo delle regioni meridionali ("Mezzogiorno").
- **TOPIC 6-Bandi:** Questo tema contiene le specifiche di una serie di bandi messi a disposizione, nell'arco dell'anno in analisi, descrivendo i dettagli e le specifiche per parteciparvi.
- **TOPIC 7 -Fondi e Incentivi:** Questo tema attira termini e concetti tecnici che riguardano la gestione dei fondi e i vari incentivi messi a disposizione dall'Unione Europea.
- **TOPIC 8 -Lavoro e Occupazione:** Questo tema gira sulla formazione dei giovani lavoratori. L'argomento affronta l'assegnazione dei fondi strutturali ai beneficiari che gestiscono i programmi di formazione volti a creare competenze professionali per i giovani.

- **TOPIC 9 -Situazione Puglia:** Questo tema descrive la situazione della regione Puglia in merito all'utilizzo di questi fondi e ne valuta gli aspetti sia positivi che negativi.
- **TOPIC 10 -Frodi:** Questo tema mette in risalto gli aspetti negativi dell'utilizzo di questi fondi, raccoglie reclami riguardanti l'inefficienza dei politici nella gestione del sistema economico italiano e, in particolare, la possibilità di assorbire i fondi UE.

Dall'analisi dei risultati emersi ed elencati in tabella possiamo notare che l'incidenza di ogni *topic* sull'intera raccolta è la seguente:

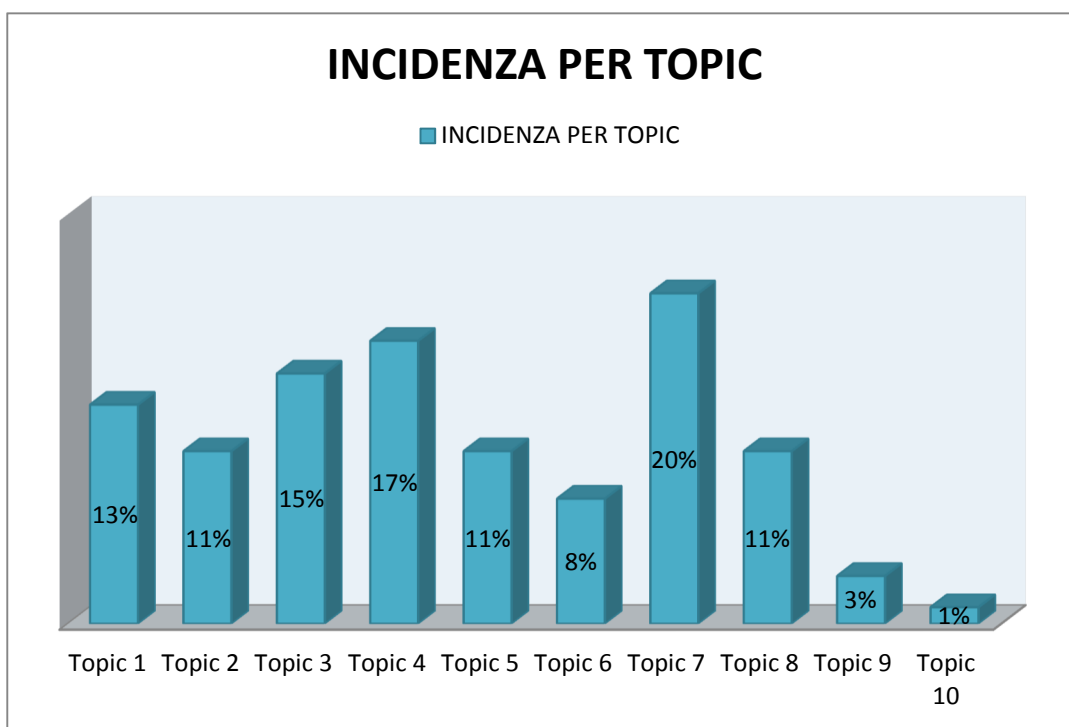


Grafico 6: Rappresentazione grafica dell'incidenza di ogni *topic* sulla raccolta di articoli, in termine di percentuale. Fonte: Ricostruzione grafica personale.

Come possiamo notare dal precedente grafico, il topic con maggiore incidenza è il topic 7 riguardante i Fondi e gli Incentivi, in effetti nella maggior parte degli articoli analizzati era inevitabile non riscontrare questi due termini che rappresentano lo scopo principale su cui è concentrata la politica di coesione.

A seguire, con il 17%, il topic riguardante le Startup e i Progetti finanziati per aumentare la crescita e lo sviluppo delle varie regioni rafforzando la ricerca e le innovazioni.

Con quasi la stessa incidenza, ovvero il 15 %, abbiamo il topic 3 che rappresenta i piani da attuare e le riforme proposte e approvate dal governo Italiano, e il topic 1, con il 13 %, che descrive i vari obiettivi della politica di coesione ridurre le disparità tra le varie regioni.

Il topic 2 e il topic 5 hanno la stessa incidenza dell'11% ed entrambi descrivono la situazione Italiana, nel secondo caso prendendo in considerazione nello specifico la situazione delle zone del Mezzogiorno.

Infine abbiamo il Topic 6, con l'8% di incidenza riguardante i Bandi messi a disposizione dall'UE, il Topic 9, con solo il 3%, che descrive la situazione specifica della Puglia e il Topic 10, che occupa una piccolissima parte della raccolta, con l'1% di incidenza che mette in risalto gli aspetti negativi dell'utilizzo di questi fondi.

Da tale analisi abbiamo valutato anche quelli che sono i pareri e le opinioni dei vari attori in merito al tema che abbiamo voluto affrontare.

Come ci mostra il seguente grafico, il 72% degli attori hanno espresso un'opinione positiva, il 20% negativa e l'8% neutra.



Grafico 7: Rappresentazione della posizione dei vari attori in merito alla raccolta analizzata, in termine di percentuale. Fonte: Ricostruzione grafica personale.

Possiamo quindi affermare che la maggior parte degli attori ha espresso un parere positivo, in quanto, il 72% sostiene che la politica di coesione contribuisca alla creazione di nuovi posti di lavoro, aumenti la competitività tra imprese, sostenga la crescita economica e lo sviluppo sostenibile e migliori la qualità di vita dei cittadini in tutte le regioni e le città dell'Unione europea.

Il 20% degli attori, invece, ha una visione abbastanza negativa in quanto, crede che nonostante le ingenti somme messe a disposizione, il divario tra le varie regioni Europee è ancora molto marcato. Il loro parere negativo non è mirato all'attuazione di tale politica ma alla scarsa capacità di gestione di tali Fondi soprattutto per quanto riguarda la gestione della spesa, i meccanismi di allocazione delle risorse e le strategie di sviluppo delle politiche.

Il restante 8% assume una posizione neutra, ovvero, non esprime un vero e proprio parere in merito all'argomento ma, solitamente, discute sui risultati ottenuti negli anni elencando gli aspetti positivi e negativi.

5.2 Analisi attraverso il *Topic Modeling*

Per quanto riguarda l'analisi *Topic Modeling*, per prima cosa abbiamo scaricato la raccolta di articoli in formato digitale tramite il sito ufficiale de *Il sole 24 Ore*. Attraverso una ricerca avanzata è possibile specificare il periodo e la parola chiave che si vuole ricercare.

Così facendo abbiamo specificato lo stesso periodo degli articoli cartacei che ci sono stati dati dalla Regione Puglia, ovvero, dal 18/03/2016 al 18/12/2017, in modo tale da poter effettuare, successivamente, un confronto con l'analisi precedentemente esposta, e specificato le parole chiave che abbiamo voluto utilizzare per tale ricerca.

I risultati ottenuti da tale ricerca sono i seguenti:

PAROLE CHIAVE	NUMERO ARTICOLI
Politica di Coesione	230
Fondi di Coesione	107
Fondi Strutturali	323
Puglia Politica di Coesione	13
Puglia Fondi di Coesione	19
Puglia Fondi Strutturali	26
Fondi Europei e Frodi	9

Tab. 2 Sono riportate nella prima colonna la lista delle parole chiave scelte per la ricerca dei documenti e nella seconda colonna il numero di documenti ricercati per ogni parola chiave.

Fonte: Ricostruzione personale.

Il sito permette, per ogni parola chiave, di selezionare tutti gli articoli insieme e scaricarli in un unico file in formato .pdf.

Così facendo abbiamo ricavato sette file, uno per ogni parola chiave contenente tutti gli articoli trovati per ognuno.

Una volta recuperati tutti gli articoli per la nostra analisi bisognava dividere singolarmente ogni articolo presente in ognuno dei sette file.

Dividere il tutto manualmente avrebbe richiesto un tempo esagerato, in quanto, come si evince dalla tabella, ci toccava creare circa 800 file, uno per ogni articolo.

Per fortuna esistono dei software semplici da utilizzare, come **TextWedge**, che permettono di dividere un file di testo in tanti piccoli file contenenti, in questo caso, ognuno un singolo articolo:

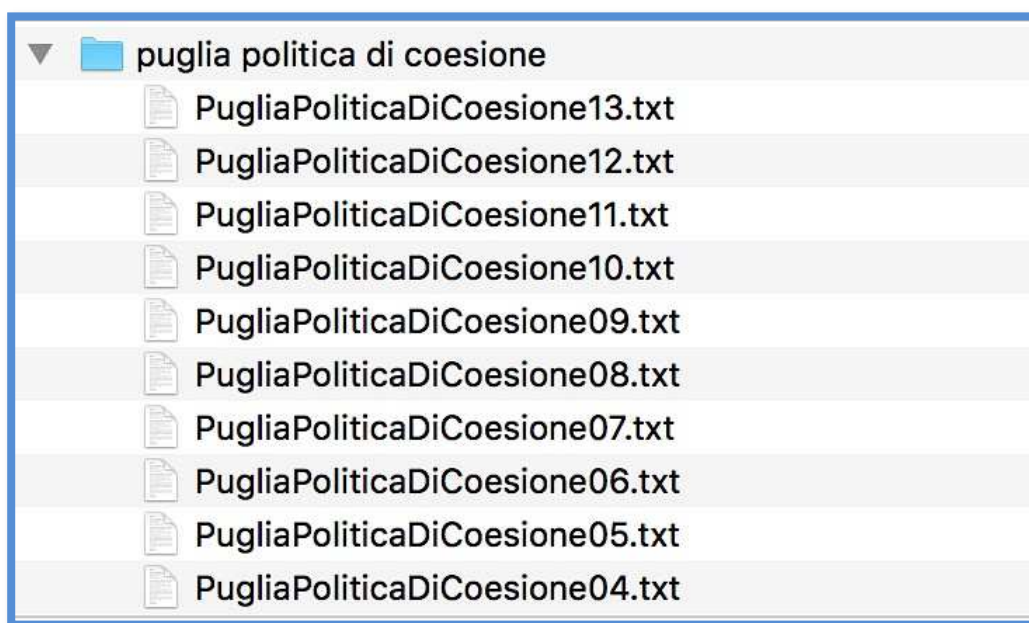


Fig.3: Rappresenta un esempio di cartella, creata tramite il software *TextWedge*, contenente la lista dei singoli articoli in formato *txt* ordinati. Fonte: Ricostruzione personale.

Sistemati tutti gli articoli bisogna scaricare il software **Mallet** per procedere con l'analisi della nostra raccolta.

Per prima cosa si deve spostare la cartella creata "Articoli" all'interno della cartella Mallet/Sample-data/web. Fatto ciò basta lanciare il primo comando Mallet per trasformare tutti gli articoli in formato .Mallet.

```
./bin/mallet import-dir --input sample-data/web/articoli --output  
articoli.mallet --keep-sequence --stoplist-file stoplist_file_ita.txt
```

Una volta creato l'output "articoli.mallet" possiamo analizzare gli articoli presenti all'interno lanciando il seguente comando:

```
./bin/mallet train-topics --input articoli.mallet --num-topics 10 --  
num-top-words 100 --num-iterations 5000 --output-state output.gz --  
output-topic-keys output-keys.txt --output-doc-topics output-  
composition.txt --diagnostics-file output.xml
```

Come spiegato nel capitolo precedente, da tale comando si ricavano quattro output diversi ovvero:

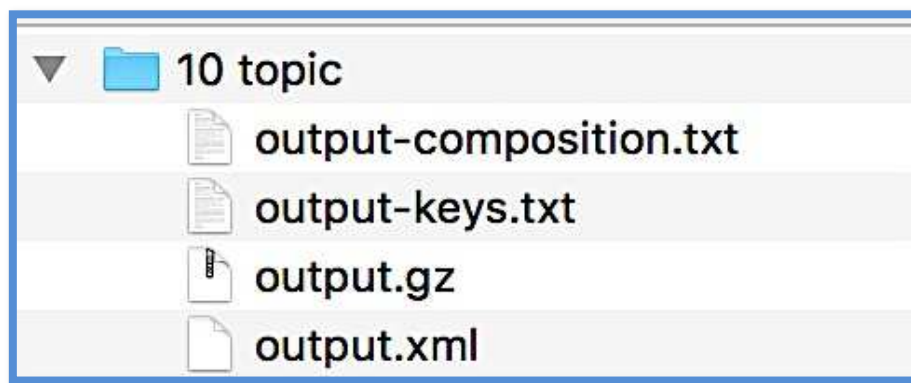


Fig.4: Rappresenta la cartella contenente tutti gli output ottenuti dal comando spiegato.

Fonte: Ricostruzione personale.

Inizialmente, si è deciso di lavorare a 15, 20 e 25 TOPIC di uguale peso, successivamente si è deciso di lavorare anche a 10 Topic, pari al numero dei temi ricavati dall'analisi qualitativa, per analizzare maggiormente le differenze tra le due tipologie di analisi.

Per comprendere meglio i risultati ottenuti si è preferito importare i due file, output-composition.txt e output-keys.txt, in un foglio Excel, così da poter leggere ed analizzare meglio i dati ottenuti.

Analizzando la lista di parole chiave restituite dal *Software Mallet*, siamo riusciti a dedurre che si sono mostrati più significative le analisi effettuate a 10 e a 15 Topic che mostreremo nel dettaglio nei prossimi paragrafo.

5.2.1 Risultati analisi a 10 Topic

Nelle seguenti tabelle saranno riportati i risultati ottenuti dall'analisi a 10 Topic suddivisi da 0-4 (Tabella 4), e da 5-9 (Tabella 4.1).

0	1	2	3	4
RISORSE A DISPOSIZIONE	LAVORO, OCCUPAZIONE E FORMAZIONE	PIANI E RIFORME DEL GOVERNO ITALIANO	MERCATI FINANZIARI	POLITICA EUROPEA
MILIONI	LAVORO	GOVERNO	CRESCITA	POLITICA
FONDI	SOCIALE	BILANCIO	BANCHE	L'EUROPA
RISORSE	CULTURALE	LEGGE	MERCATI	PAESE
MILIARDI	TILOCBLOB	ANNI	PAESI	PAESI
REGIONI	LAVORATORI	LAVORO	DEBITO	POLITICO
SVILUPPO	CULTURALI	MESI	MERCATO	LEADER
PROGETTI	PROFESSIONISTI	GIÀ	RISCHIO	GRANDE
EURO	IMPRESE	MISURE	BCE	GOVERNO
SPESA	GESTIONE	CONTI	BANCA	GERMANIA
SUD	ATTIVITÀ	EURO	INVESTIMENTI	ANNI
PIANO	SETTORE	MILIONI	CRISI	PARTITO
COESIONE	BENI	MILIARDI	POLITICA	ELEZIONI
INVESTIMENTI	CULTURA	GIORNI	TASSI	COSÌ
IMPRESE	FORMAZIONE	MINISTRO	SISTEMA	MIGRANTI
EUROPEI	SERVIZI	ANNO	CAPITALE	REFERENDUM
INTERVENTI	PATRIMONIO	TEMPO	EUROPEA	CRISI
PROGRAMMI	PARTECIPAZIONE	ROMA	STABILITÀ	TEMPO
NAZIONALE	CASO	MANOVRA	RIFORME	EUROPA
PROGRAMMA	ATTRAVERSO	GIOVANI	PUÒ	VUOLE
GIÀ	SVILUPPO	RENZI	PUBBLICO	PARTITI
STRUTTURALI	PUBBLICA	RIFORMA	ECONOMICA	VOTO
PROGRAMMAZIONE	SOCIETÀ	NUOVO	POSSONO	CONTRO
FINANZIAMENTI	ECONOMICO	CASO	ANNI	FINE
MEZZOGIORNO	FORMA	COMMISSIONE	TITOLI	STABILITÀ
AREE	CASSA	RICOSTRUZIONE	MONETARIA	FUTURO
TOTALE	FAVORIRE	INCENTIVI	LIVELLO	GRANDI
REGIONALI	SOCIALI	MISURA	BANCARIO	GOVERNI
MINISTERO	NUOVA	POTREBBE	RIPRESA	PRESIDENTE
PMI	PUÒ	FRONTE	L'ECONOMIA	COMUNE
CREDITO	INOLTRE	DECRETO	PIL	MERKEL

Tab.4: Tabella contenente i primi 5 Topic con la lista di parole chiave scelte per ognuno dal Software Mallet.

5	6	7	8	9
POLITICHE STRUTTURALI	POLITICHE PUBBLICHE E INTERVENTI INFRASTRUTTURALI	EFFETTO ECONOMICO FONDI STRUTTURALI SU RICERCA E IMPRESA	SITUAZIONE PUGLIA	POLITICA INDUSTRIALE NEL MEZZOGIORNO
FONDI	MILIARDI	RICERCA	MILIONI	PRESIDENTE
EUROPEA	INVESTIMENTI	ANNI	REGIONE	PAESE
EUROPEO	INFRASTRUTTURE	EURO	PUGLIA	IMPRESE
PAESI	CITTÀ	AZIENDE	PATTO	CRESCITA
STRUTTURALI	GOVERNO	IMPRESE	TURISMO	CONFINDUSTRIA
L'ITALIA	PUBBLICI	MERCATO	MILIARDI	POLITICA
BRUXELLES	OPERE	GRANDI	PRESIDENTE	BOCCIA
EUROPEI	PIANO	SISTEMA	GOVERNATORE	SUD
COMMISSIONE	INTERVENTI	SOCIETÀ	SVILUPPO	INVESTIMENTI
COESIONE	CIPE	MODELLO	C'È	POLITICHE
POLITICA	SVILUPPO	CRISI	REGIONALE	SVILUPPO
POLITICHE	PATTI	MONDO	EMILIANO	INDUSTRIALE
MILIARDI	PROGETTI	MILANO	GOVERNO	PUÒ
BILANCIO	LOCALI	UNIVERSITÀ	LAVORO	SISTEMA
C'È	SERVIZI	INDUSTRIALE	PROGRAMMA	GRANDE
ITALIA	PUNTO	CITTÀ	INFRASTRUTTURE	COMPETITIVITÀ
GERMANIA	SPESA	NUOVO	FONDI	CENTRO
DELL'UNIONE	MINISTRO	SPIEGA	COOPERAZIONE	GOVERNO
FRANCIA	PATTO	ITALIA	COMUNE	RUOLO
MEMBRI	FSC	QUALITÀ	GIÀ	PROGETTO
PAESE	MESI	INNOVAZIONE	TARANTO	DOBBIAMO
SPAGNA	C'È	ITALIANO	AMBIENTALE	LAVORO
PIL	PRIORITÀ	VALORE	PERIODO	C'È
RISORSE	ANNI	TECNOLOGIA	IMPRESE	QUESTIONE
PROPOSTA	METROPOLITANE	POCO	MESI	GIOVANI
REGIONI	PUBBLICA	NUOVI	INTERVENTI	NAZIONALE
EUROPEE	COMUNI	DIGITALE	RISORSE	ECONOMICA
GIÀ	FONDI	PRODUZIONE	COESIONE	COESIONE
POLONIA	PERÒ	TECNOLOGIE	TERRITORIO	INDUSTRIA
RISERVATA	ENTRO	ATTRAVERSO	SINDACO	DARE

Tab.4.1: Tabella contenente i Topic da 5 a 9, con la lista di parole chiave scelte per ognuno dal Software Mallet. Fonte: Ricostruzione personale

Una volta analizzate le singole parole chiave, siamo riusciti a dare un titolo ad ogni singolo *Topic* che ci permettesse di comprendere al meglio il tema più discusso in merito alla Politica di Coesione nell'anno prestabilito.

Possiamo riassumere gli argomenti in questo modo:

- **TOPIC 0 – Risorse a disposizione:** Questo *Topic* descrive come le politiche di coesione impiegano risorse pubbliche, europee e nazionali, con la finalità di rimuovere le disuguaglianze di sviluppo, incrementare le opportunità di crescita e inclusione sociale dei cittadini e promuovere la coesione economica fra i territori italiani.
- **TOPIC 2 – Lavoro, Occupazione e Formazione:** Questo *Topic* gira sulla formazione di posti di lavoro. In effetti, “lavoro”, “formazione” e “sviluppo” sono le tre parole che caratterizzano il tema. Tra le seguenti parole, il concetto di formazione professionale si ripete. Il *Topic* affronta l'assegnazione dei fondi strutturali ai beneficiari che gestiscono i programmi di formazione volti a creare competenze professionali.
- **TOPIC 2 – Piani e riforme del Governo Italiano:** Questo *Topic* descrive tutti i piani adottati e le leggi approvate dal governo Italiano per quanto riguarda la Politica di Coesione. In effetti, le parole che caratterizzano il tema sono “governo”, “legge”, “Renzi” e “riforma”.
- **TOPIC 3 – Mercati Finanziari:** Questo *Topic* cattura il controverso dibattito riguardante il ruolo dell'Europa nella gestione dei sistemi finanziari. Il tema si riferisce specificamente alle “banche”, “debiti”, “capitale”.

- **TOPIC 4 –Politica Europea:** Questo *Topic* descrive le politiche Europee che dovrebbero mirare e ad aumentare l'unità europea soprattutto in ambito sociale, economico ed istituzionale.
- **TOPIC 5 –Politiche Strutturali:** Questo *Topic* descrive l'insieme delle azioni strutturali dirette a ridurre i livelli di sottosviluppo tra le regioni e gli stati membri dell'UE.
- **TOPIC 6 –Politiche Pubbliche e interventi Infrastrutturali:** Questo *Topic* descrive i vari interventi e investimenti rivolti alle infrastrutture della città. Infatti, parole come “infrastruttura”, “metropolitane”, “servizi”, sono associate a “programmazione”, “priorità”, “interventi”.
- **TOPIC 7 – Effetto Economico Fondi Strutturali su ricerca e impresa:** Questo *Topic* descrive gli effetti positivi e negativi inerenti all'utilizzo di questi fondi per la crescita del mercato italiano, su questioni regionali specifiche riguardanti le imprese, le società e le varie innovazioni tecnologiche come si può notare dalle parole “mercato”, “aziende”, “tecnologie”, “Innovazione”.
- **TOPIC 8 –Situazione Puglia:** Questo *Topic* descrive la situazione specifica della regione puglia in merito all'utilizzo di questi fondi e dei vari interventi sul territorio con maggior interesse nell'ambito del turismo.
- **TOPIC 9 –Politica Industriale nel Mezzogiorno:** Questo tema si concentra sul rapporto tra fondi strutturali e lo sviluppo industriale delle regioni meridionali Italiane (“Mezzogiorno”). Parole come, “crescita” “impresa” sono associate a “investimenti” e “sviluppo”.

5.2.2 Risultati analisi a 15 Topic

Nelle seguenti tabelle saranno riportati i risultati ottenuti dall'analisi a 15 Topic suddivisi da 0-4 (Tabella 5), da 5-9 (Tabella 5.1), e da 10-14 (Tabella 5.2).

0	1	2	3	4
POLITICA EUROPEA	RUMORE	ANALISI DELLE QUESTIONI DI BILANCIO	FONDI STRUTTURALI PER IMPRENDITORIALITA' CREATIVA	PIANI E RIFORME DEL GOVERNO ITALIANO
POLITICA	ANNI	CRESCITA	SOCIALE	GOVERNO
EUROPEA	GRANDE	INVESTIMENTI	CULTURALE	MINISTRO
PAESI	STORIA	BILANCIO	SETTORE	ROMA
L'EUROPA	SOCIETÀ	MILIARDI	CULTURALI	RENZI
POLITICHE	PAESE	GOVERNO	POLITICA	GIÀ
GERMANIA	CITTÀ	PIL	BENI	GIORNI
EUROPEO	COSÌ	ANNI	SVILUPPO	PRESIDENTE
EUROPA	SENSO	PUBBLICI	CULTURA	RICOSTRUZIONE
PAESE	SOCIALE	SPESA	PATRIMONIO	CONSIGLIO
L'ITALIA	MONDO	MISURE	IMPRESE	COMUNI
POLITICO	VALORI	PUBBLICO	PUÒ	BILANCIO
PRESIDENTE	CASA	FISCALE	ATTRAVERSO	PROCEDURE
L'UNIONE	REPUBBLICA	DELL'ECONOMIA	ATTIVITÀ	COMMISSIONE
LEADER	GRANDI	DEBITO	ECONOMICA	LEGGE
NAZIONALI	TEMPO	PERÒ	PUBBLICA	PALAZZO
EUROPEI	PERSONE	STRUTTURALI	ECONOMICO	TERREMOTO
ELEZIONI	POLITICA	DATI	PRODUZIONE	PREMIER
STABILITÀ	DEMOCRAZIA	RISORSE	NUOVE	RIPRODUZIONE
RIFORME	COESIONE	COSÌ	STRUMENTI	MOMENTO
BREXIT	L'ECONOMIA	DEFICIT	SERVIZI	RISCHIO
GOVERNI	MOLTI	FISCALI	GESTIONE	MINISTRI
DRAGHI	COMUNE	ANNO	POSSONO	PROPOSTA
ECONOMICA	PROBLEMA	LAVORO	PRODOTTI	MILIONI
PARTITO	GIORNO	PUBBLICA	TUTELA	MIGRANTI
PROGETTO	TANTI	ECONOMICA	SOSTEGNO	RISERVATA
RISERVATA	POTERE	RIPRESA	FONDAZIONE	NUOVO
PUÒ	PESCATORE	STABILITÀ	CENTRI	POSSIBILE
DELL'UNIONE	IMMIGRATI	L'ITALIA	CICLO	PRESIDENZA
GRANDE	GUERRA	RIDUZIONE	CREATIVA	GENTILONI
BCE	DIRITTO	RAPPORTO	SOSTENIBILITÀ	LAVORO

Tab.5: Tabella contenente i 15 Topic con la lista di parole chiavi scelte per ognuno dal Software Mallet

5	6	7	8	9
SITUAZIONE EUROPA	MERCATI FINANZIARI	RISORSE A DISPOSIZIONE	POLITICHE STRUTTURALI	POLITICA INDUSTRIALE
TILOCBLOB	BANCHE	MILIONI	FONDI	IMPRESE
PAESI	MERCATI	PROGETTI	COESIONE	PRESIDENTE
SPAGNA	BANCA	RISORSE	REGIONI	CONFINDUSTRIA
FRANCIA	RISCHIO	MILIARDI	MILIARDI	PAESE
POLONIA	MERCATO	PIANO	STRUTTURALI	BOCCIA
BRUXELLES	TASSI	FONDI	RISORSE	INDUSTRIALE
GERMANIA	PAESI	SVILUPPO	COMMISSIONE	RICERCA
BRETAGNA	TITOLI	REGIONI	EUROPEI	SVILUPPO
COOPERAZIONE	CRESCITA	EURO	BRUXELLES	CENTRO
FONDI	BANCARIO	IMPRESE	POLITICHE	CRESCITA
ORBAN	CAPITALE	PROGRAMMI	BILANCIO	COMPETITIVITÀ
REPUBBLICA	INVESTITORI	PROGRAMMA	L'ITALIA	POLITICHE
DELL'EST	BCE	GIÀ	MILIONI	GIOVANI
COESIONE	FINANZIARI	INTERVENTI	EUROPEO	INVESTIMENTI
LONDRA	MONDIALE	FINANZIAMENTI	POLITICA	DOBBIAMO
CECA	GLOBALE	INVESTIMENTI	EUROPEA	QUESTIONE
MILIARDI	PROBLEMA	INFRASTRUTTURE	EURO	MONDO
EURO	CRISI	PMI	MEMBRI	LAVORO
PARTNER	RISCHI	ENTRO	NAZIONALE	INDUSTRIALI
REGNO	ITALIA	AREE	REGIONALI	NAZIONALE
GRAN	CINA	MINISTERO	PAESI	AZIENDE
ITALIANE	BASSI	EUROPEI	INVESTIMENTI	GRANDE
SPIEGA	CREDITO	BANDO	ITALIA	INDUSTRIA
EUROPEI	FONDI	RICERCA	SPESA	FORTE
UNGHERIA	MONDO	EUROPEA	PROPOSTA	PIANO
MIGRANTI	ANNI	PROGETTO	PAESE	TERRITORIO
POSTO	LIVELLO	NAZIONALE	RISORSE	ITALIA
ECONOMIE	STRUTTURALI	NAZIONALI	PARI	POLITICA
SEGUITA	DOLLARI	BANDI	DELL'UNIONE	FUTURO
GRUPPO	MONETARIA	STRUTTURALI	RISERVATA	INNOVAZIONE

Tab.5.1: Tabella contenente i 15 Topics con la lista di parole chiave scelte per ognuno dal *Software Mallet*.

10	11	12	13	14
RUMORE	LAVORO, OCCUPAZIONE E FORMAZIONE	SITUAZIONE MEZZOGIORNO	SITUAZIONE PUGLIA	POLITICHE PUBBLICHE E INTERVENTI INFRASTRUTTURALI
sistema	lavoro	sud	puglia	miliardi
nuovo	legge	milioni	regione	investimenti
riforma	euro	miliardi	milioni	città
può	giovani	mezzogiorno	regionale	opere
anni	professionisti	coesione	sviluppo	governo
punto	lavoratori	risorse	miliardi	cipe
potrebbe	mesi	fondi	imprese	pubblici
ciò	imprese	investimenti	crescita	locali
caso	contratti	sviluppo	emiliano	mesi
già	formazione	spesa	risorse	sviluppo
ultimi	anni	interventi	investimenti	infrastrutture
nuova	caso	Campania	governo	metropolitane
futuro	tempo	regioni	euro	fsc
fine	massimo	patti	presidente	fondi
regole	bonus	già	ambientale	programmazione
fronte	fondi	vincenti	lavoro	delirio
crisi	spese	Calabria	sistema	punto
nuovi	misura	patto	anni	servizi
processo	contratto	decreto	c'è	patto
fiducia	incentivi	governo	patto	piano
quadro	strutturali	Sardegna	fondi	priorità
base	gennaio	anni	sociale	efficienza
tempo	contributi	infrastrutture	politica	anni
fase	europei	ministro	mesi	obiettivi
tempi	misure	Sicilia	economico	ministro
particolare	forma	Basilicata	periodo	c'è
ruolo	norma	credito	territorio	interventi
realità	autonomi	Roma	turismo	spesa
grado	garanzia	nazionale	Taranto	Milano
così	partecipazione	puglia	può	decreto

Tab.5.2: Tabella contenente i 15 Topic con la lista di parole chiave scelte per ognuno dal *Software Mallet*. Fonte: ricostruzione personale

Anche in questo caso abbiamo analizzato i risultati emersi ed associato, ad ogni singolo Topic, un titolo:

- **TOPIC 0 –Politica Europea:** Questo Topic descrive le politiche Europee che dovrebbero mirare e ad aumentare l'unità europea soprattutto in ambito sociale, economico ed istituzionale.
- **TOPIC 1/10 –“Rumore”:** è l’algoritmo del TM che mette insieme parole poco significative (come “ciò”, “può”, “già”), al fine di aumentare la coerenza degli altri *Topic*.
- **TOPIC 2 –Analisi delle questioni di Bilancio:** Questo Topic descrive argomenti relativi alla situazione finanziaria, nelle sue sfaccettature di reddito e di spese. Analisi della tassazione, sistema fiscale, deficit di bilancio, sistema fiscale.
- **TOPIC 3 –Fondi Strutturali Per Imprenditorialità Creativa:** Questo *Topic* collega i fondi strutturali con l'imprenditorialità culturale regionale. Si evince dalla presenza delle parole “culturali”, “beni”, “patrimonio”.
- **TOPIC 4 –Piani e Riforme del Governo Italiano:** Questo *Topic* descrive tutti i piani adottati e le leggi approvate dal governo Italiano per quanto riguarda la Politica di Coesione. In effetti, le parole che caratterizzano il tema sono “governo”, “legge”, “Renzi” e “riforma”.
- **TOPIC 5 – Situazione Europa:** Questo *Topic* descrive la situazione Europea in merito all’utilizzo di questi fondi e dei

vare interventi sul territorio per ridurre il divario tra le varie regioni Europee.

- **TOPIC 6 –Mercati Finanziari:** Questo *Topic* cattura il controverso dibattito riguardante il ruolo dell'Europa nella gestione dei sistemi finanziari. Il tema si riferisce specificamente alle "banche", "debiti", "capitale".
- **TOPIC 7 –Risorse a disposizione:** Questo *Topic* descrive come le politiche di coesione impiegano risorse pubbliche, europee e nazionali, con la finalità di rimuovere le disuguaglianze di sviluppo, incrementare le opportunità di crescita e inclusione sociale dei cittadini e promuovere la coesione economica fra i territori italiani.
- **TOPIC 8 –Politiche Strutturali:** Questo *Topic* descrive l'insieme delle azioni strutturali dirette a ridurre i livelli di sottosviluppo tra le regioni e gli stati membri dell'UE.
- **TOPIC 9 –Politica Industriale:** Questo *Topic* racchiude l'insieme di misure a sostegno del settore industriale o secondario adottate dall'UE per rafforzare la competitività del settore manifatturiero europeo e di conseguenza l'intera economia europea.
- **TOPIC 11 – Lavoro, Occupazione e Formazione:** Questo *Topic* gira sulla formazione dei giovani lavoratori. In effetti, "lavoro", "giovani" e "occupazione" sono le tre parole che caratterizzano il tema. Tra le seguenti parole, il concetto di formazione professionale si ripete. L'argomento probabilmente affronta l'assegnazione dei fondi strutturali ai beneficiari che gestiscono i programmi di formazione volti a creare competenze

professionali per i giovani. L'idea che emerge è che questa attività potrebbe forse aiutare i giovani a diventare impiegato.

- **TOPIC 12 –Situazione Mezzogiorno:** Questo Topic si concentra sul rapporto tra fondi strutturali e lo sviluppo delle regioni meridionali (“Mezzogiorno”). Lo notiamo dalle Parole come “risorse” “interventi” “vincenti” sono associati a “investimenti”.

- **TOPIC 13 –Situazione Puglia:** Questo Topic descrive la situazione specifica della regione puglia in merito all’utilizzo di questi fondi e dei vari interventi sul territorio con maggior interesse nell’ambito del turismo.

- **TOPIC 14 - Politiche Pubbliche e interventi Infrastrutturali:** Questo Topic descrive i vari interventi e investimenti rivolti alle infrastrutture della città. Infatti, parole come “infrastruttura”, “metropolitane”, “servizi”, sono associati a “programmazione”, “priorità”, “interventi”.

5.3 Confronto tra i risultati ottenuti con l'analisi *Topic Modeling*

Come possiamo notare dal confronto delle due analisi fatte a 10 e a 15 Topic, molti dei topic analizzati sono risultati simili, se non identici, tra loro:

- **Topic 4 con Topic 0 –Politiche Europee:** più della metà delle parole emerse sono identiche tra i due topic e le restanti comunque restano simili o inerenti al tema.
- **Topic 2 con Topic 4 - Piani e Riforme del Governo Italiano:** quasi la metà delle parole sono uguali le restanti sono comunque simili.
- **Topic 6 con Topic 14: Politiche Pubbliche e Interventi Infrastrutturali:** La maggior parte delle parole sono identiche. Lo stesso avviene tra i topic riguardanti i **Mercati Finanziari** (Topic 3 e 6), **Risorse a disposizione** (Topic 0 e 7), **Politiche Strutturali** (Topic 5 e 8), **Situazione Puglia** (Topic 8 e 13).
- **Topic 1 con Topic 11 - Lavoro, Occupazione e Formazione:** In questo caso sono poche le parole che risultano uguali tra i due topic, quasi una decina, ma il tema affrontato resta il lavoro. Questo perché, analizzando bene le parole, si evince che nel topic a 15 il tema è più incentrato sulla formazione dei giovani disoccupati mentre nel Topic a 10 si parla di lavoro e occupazione in generale.

La stessa osservazione viene fatta tra i topic inerenti alla **Situazione del Mezzogiorno**, in quanto, il topic dell'analisi a 10 è incentrato nello specifico sul tema della politica industriale per le regioni del mezzogiorno mentre nell'analisi a 15 sugli investimenti e gli interventi.

Detto ciò possiamo affermare che sono poche le differenze riscontrate tra l'analisi a 10 Topic e l'analisi a 15. Solo il topic inerente all' "*Effetto Economico Fondi Strutturali su Ricerca e Impresa*" non rientra nell'analisi a 15. Inoltre due dei Topic a 15 (1 e 10) sono classificati come Topic "Rumore", ovvero contengono una lista di parole non significative per far sì che si aumenti la coerenza tra gli altri Topic.

Ciò non vuole dire che l'analisi a 15 è stata inutile, anzi, ci ha permesso di trovare altri **3 topic** molto significativi e più specifici, ovvero i topic riguardanti:

1. *Analisi delle questioni di Bilancio,*
2. *Politica Industriale*
3. *Fondi Strutturali per Imprenditorialità Creativa.*

5.4 Confronto dei risultati tra *Analisi Qualitativa* e *Topic Modeling*

Ora che i modelli sono stati analizzati sia dal punto di vista teorico che dal punto di vista delle procedure di stima, è interessante un confronto per capire se il loro comportamento è simile o, al contrario, uno prevale sull'altro. Inizialmente abbiamo effettuato un confronto con l'analisi a 10 topic, dove possiamo subito notare che tre su dieci topic risultano affrontare lo stesso argomento ovvero:

TITOLO TOPIC	ANALISI QUALITATIVA	TOPIC MODELING	PAROLE IN COMUNE
Piani e Riforme	Topic 3	Topic 2	GOVERNO, RIFORME, LEGGE, DECRETO, INCENTIVI, MISURE, MANOVRA.
Lavoro e Occupazione	Topic 8	Topic 1	LAVORO, ATTIVITÀ, FORMAZIONE, SERVIZI, SETTORE, SOCIALE.
Situazione Puglia	Topic 9	Topic 8	PUGLIA, REGIONE, TURISMO, INTERVENTI, RISORSE, SVILUPPO, INFRASTRUTTURE.

Tab.6: tabella contenente i tre topic e le corrispettive parole chiave, emersi in entrambe le due tipologie di analisi. Fonte: Ricostruzione personale.

Osservando il primo dei tre topic in tabella e analizzando la lista delle parole chiave che sono emerse dalle due analisi possiamo affermare che tutti gli articoli che appartengono a tale Topic parlano di interventi, riforme, leggi e decreti approvati e valutati dal governo italiano e delle misure e delle manovre utilizzata per sfruttare gli incentivi che ci sono messi a disposizione dall'UE.

Risulta inevitabile quindi, non notare la similarità tra i due topic a cui è stato, appunto, assegnato lo stesso titolo.

Nel caso del topic “Lavoro e Occupazione” come possiamo notare le parole chiave sono tutte volte al lavoro, alla formazione allo sviluppo e alle attività, dall’analisi qualitativa si è spesso parlato, per quanto riguarda gli articoli appartenenti a questo topic, dell’assegnazione dei fondi strutturali ai beneficiari che gestiscono i programmi di formazione volti a creare competenze professionali e, soprattutto, di garanzie di posti di lavoro rivolti ai giovani disoccupati e ai “Neet”.

Proprio per questo motivo, mi aspettavo di riscontrare dall’analisi di *Topic Modeling* tante altre parole che sono risultate, invece, molto significative nell’analisi qualitativa.

Ero abbastanza certa di riscontrare le parole “giovani”, “disoccupati”, “assunzioni” all’interno della lista ottenuta dall’analisi di *Topic Modeling*.

Infine, per quanto riguarda il topic “Situazione Puglia”, ero più che sicura di ritrovarlo in entrambe le tipologie di analisi dato che, per l’analisi qualitativa, gli articoli sono stati selezionati dalla regione Puglia, quindi era inevitabile non ritrovare un topic specifico per la Puglia; per l’analisi di *Topic Modeling*, data la scelta delle parole chiave utilizzate per effettuare la ricerca degli articoli, ovvero 3 delle 7 parole erano rivolta alla regione Puglia, era più che ovvio che almeno uno dei topic emersi rappresentasse la situazione della Puglia.

Vorrei, inoltre, sottolineare che, la maggior parte degli articoli di questo topic, descrivono la situazione della puglia in merito soprattutto alla crescita del turismo. Si è parlato spesso nei vari articoli del piano strategico regionale *Puglia 365* in cui vengono individuate sei macro-aree di intervento: prodotto, accoglienza, formazione, infrastrutture, innovazione, e promozione.

La Puglia del turismo, infatti, ha conquistato negli ultimi anni un ottimo posizionamento internazionale e questo viene colto a pieno in entrambe le tipologie di analisi applicate.

ANALISI QUALITATIVA	TOPIC MODELING	PAROLE IN COMUNE
Situazione Mezzogiorno (Topic 5)	Politica Industriale nel Mezzogiorno (Topic 9)	SUD, IMPRESA, CRESCITA, INVESTIMENTI COMPETITIVITÀ
Fondi e Incentivi (Topic 7)	Politiche Strutturali (Topic 5)	FONDI, BRUXELLES, MILIARDI, ITALIA, GERMANIA, FRANCIA.

Tab.7: tabella contenente i topic emersi dall'analisi qualitativa e dall'analisi Topic Modeling che risultano simili anche se non gli è stato assegnato lo stesso titolo e le rispettive parole chiave trovate in entrambe le analisi. Fonte: Ricostruzione personale.

Anche se con titoli diversi, molti topic risultano comunque simili.

Come possiamo notare dalla Tabella.7 il topic 5 emerso dall'analisi qualitativa è molto simile al Topic 9 dell'analisi *Topic Modeling*. Entrambi descrivono la situazione delle zone del Mezzogiorno in merito all'utilizzo di tali fondi, con la differenza che il topic 9 è improntato sul rapporto tra fondi strutturali e lo sviluppo industriale nelle regioni meridionali Italiane.

Ciò non toglie che, tale tema non sia stato riscontrato anche durante l'analisi qualitativa, anzi spesso si è parlato del *Rapporto Svimez* (associazione per lo sviluppo dell'industria nel Mezzogiorno).

È molto interessante analizzare tale topic, in quanto, le opinioni in merito riscontrate dall'analisi qualitativa erano molto contrastanti.

Tra gli attori che affrontano tale argomento c'è chi crede che l'effetto dei fondi UE ha portato ad una crescita per il Sud, altri invece credono che tale crescita è solo un miraggio e che l'utilizzo di tali fondi in queste zone è inadatto.

In entrambe le analisi sono state riscontrate le parole: sud, impresa, crescita, investimenti e competitività, ma in nessuna delle parole emerse dall'analisi Topic Modeling lascia pensare a degli aspetti negativi in merito all'argomento. Per quanto riguarda i topic "Fondi e Incentivi" e "Politiche strutturali", sostanzialmente i temi sono molto simili.

Ambedue affrontano argomenti inerenti ai diversi fondi messi a disposizione dall'UE per le varie regioni e come vengono utilizzati dalle stesse. Infatti le parole ritrovate sono: fondi, Bruxelles, miliardi, Italia, Germania, Francia.

Dopo quest'attenta analisi la curiosità per noi è stata valutare se i restanti topic emersi dall'analisi qualitativa potessero emergere aumentando il numero di topic nell'analisi *Topic Modeling* (da qui l'idea di analizzare i topic a 15). Così facendo abbiamo confrontato l'analisi qualitativa con l'analisi a 15 Topic. Abbiamo notato che gli ulteriori topic trovati iniziassero ad affrontare temi sempre più dettagliati.

È possibile notare come quattro dei 15 topic risultano identici sia a quelli trovati nell'analisi qualitativa che in quella a 10 topic:

TITOLO TOPIC	ANALISI QUALITATIVA	ANALISI TOPIC MODELING 10 TOPIC	ANALISI TOPIC MODELING 15 TOPIC
Piani e riforme	Topic 3	Topic 2	Topic 4
Lavoro, Occupazione e formazione	Topic 8	Topic 1	Topic 11
Situazione Puglia	Topic 9	Topic 8	Topic 13
Situazione Mezzogiorno	Topic 5	Topic 9	Topic 12

Tab.8: tabella contenente il nome e il numero specifico dei topic, emersi dalle tre tipologie di analisi, che risultano uguali. Fonte: Ricostruzione pers.

Analizzando le parole emerse nei topic a 15 possiamo notare, però, alcune piccole differenze che ci mostrano come tale analisi diventa più dettagliata a differenza dell'analisi a 10 topic.

Per quanto riguarda, per esempio, il topic inerente al lavoro e all'occupazione (Topic 11) che emerge sia nell'analisi qualitativa che nell'analisi a 10 topic, emergono le parole che ci aspettavamo già nel topic a 10, ovvero: giovani, contratti, bonus e garanzia.

Tali parole ci fanno capire che questo topic descrive nel dettaglio ciò che abbiamo riscontrato nell'analisi qualitativa inerente alla situazione dei giovani disoccupati e agli incentivi messi a disposizione per garantire nuovi contratti. La stessa differenza si può notare per quanto riguarda il topic che descrive la situazione delle zone del Mezzogiorno, in quanto, nell'analisi a 10 tale topic rappresenta la situazione specifica delle politiche industriali del Mezzogiorno come possiamo notare dalle parole Confindustria, industriale, politica e industria che non ritroviamo nel topic a 15. Invece, in quest'ultimo ritroviamo le parole che rappresentano la situazione generale delle regioni del Mezzogiorno, infatti troviamo parole come Campania, Calabria, Sicilia, Puglia, risorse.

Quindi possiamo confermare che il topic a 15 riferito alla situazione del Mezzogiorno è molto più vicino al topic emerso dall'analisi qualitativa seppure in alcuni articoli di quest'ultima si riscontra qualche opinione inerente alle politiche industriali del Mezzogiorno trovati nel Topic dell'analisi a 10.

Per quanto riguarda gli altri topic dell'analisi a 15, analizzando bene le parole di ognuno, abbiamo notato che alcuni di essi possono essere raggruppati in un unico topic più generico emerso dall'analisi qualitativa:

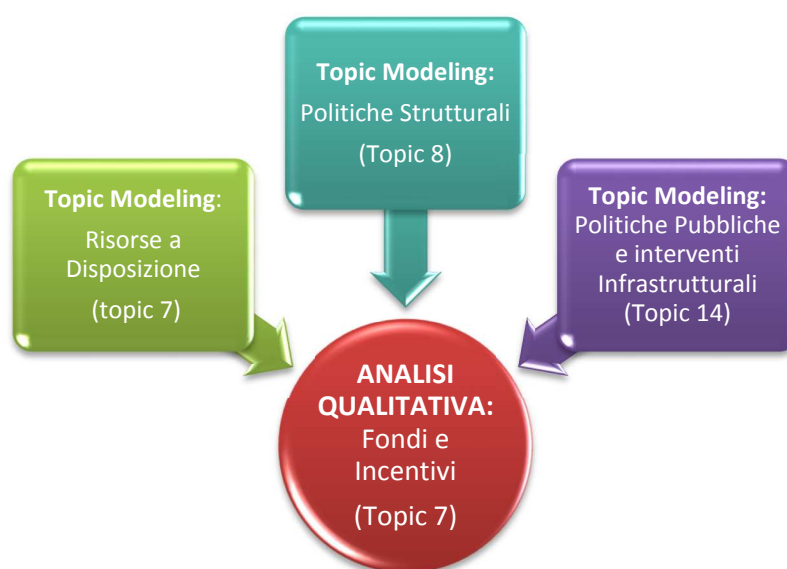


Grafico 8: Rappresentazione grafica del raggruppamento dei tre topic emersi dall'analisi Topic Modeling a 15 in un unico topic emerso nell'analisi qualitativa. Fonte: Ricostruzione grafica personale.

Come possiamo notare dal nostro grafico, il Topic 7 (Fondi e Incentivi) va a raggruppare i topic 7, 8 e 14 emersi dall'analisi *Topic Modeling*. Questo perché, in ognuno di questi topic presenta parole in comune, come “fondi”, “Investimenti”, “Risorse”, “Miliardi”, “Incentivi”.

Ognuno di questi topic affronta il tema inerente alla gestione dei fondi e delle risorse messe a disposizione dall'Unione Europea per rimuovere le disuguaglianze tra le regioni e gli stati membri dell'UE e incrementare le opportunità di crescita attraverso i vari interventi.

Uno dei topic che mi ha colpito maggiormente è stato il topic 4 risultato dall'analisi qualitativa, “Startup e Progetti”. Ero più che convinta di ritrovare lo stesso topic anche nell'analisi *Topic Modeling* in quanto nell'analisi qualitativa questo topic ha una maggiore incidenza sulla nostra raccolta, addirittura copre il secondo posto con il 17%. Si è parlato addirittura negli ultimi anni di crescita record per l'Italia in merito alle Startup, soprattutto nel nord Italia. Molti sono stati gli articoli che hanno affrontato tale tema ma, dall'analisi di Topic non c'è traccia della parola “Startup” in nessun topic.

Per concludere, andando ad analizzare ed a confrontare l'incidenza dei vari topic sulla nostra raccolta, possiamo affermare di aver ottenuto lo stesso risultato tra l'analisi qualitativa e l'analisi *Topic Modeling* a 15. Come abbiamo visto nel Grafico.6 il topic con maggiore incidenza nell'analisi qualitativa è il topic 7 (Fondi e Incentivi).

Dato che, la maggior parte dei topic, come abbiamo visto nel grafico 9, contengono articoli che parlano dei vari tipi di fondi messi a disposizione dall'Unione Europea, di come vengono utilizzati dalle varie regioni, dell'importanza che assumono per la crescita del loro bilancio, possiamo dire che anche nell'analisi *Topic Modeling* l'argomento con maggiore incidenza tratta di Fondi e Incentivi.

CONCLUSIONI

Lo scopo del nostro studio era dimostrare come l'analisi di un testo è importante per riuscire a comprendere i pensieri della società in merito a qualsiasi argomento.

Non consiste solamente nella semplice estrazione delle parole identificative di un argomento, ma bensì della classificazione dei documenti in base ai temi trattati, nella speranza di migliorarne l'archiviazione, la ricerca e l'esplorazione.

Per far ciò ci siamo concentrati su due tipologie di analisi che ci permettono di ottenere tali risultati: l'analisi qualitativa e l'analisi *Topic Modeling*.

Lo scopo della tesi è stato quello di esplorare, analizzare e confrontare i due metodi scelti per analizzare una raccolta di dati focalizzandoci soprattutto sul confronto dei risultati ottenuti.

Prima di iniziare ad analizzare un qualsiasi testo è di fondamentale importanza avere una buona conoscenza dei dati che si vogliono analizzare per scoprire quali dei topic estratti caratterizzano al meglio un determinato argomento e in che modo lo affrontano.

L'argomento specifico che abbiamo scelto di trattare in questo elaborato è la "Politica di Coesione". Attraverso una serie di articoli, cartacei e digitali, del quotidiano Italiano *Il Sole 24 Ore*, che coprono il periodo di tempo che va dal 18/03/2016 al 18/12/2017, abbiamo applicato tali metodi e confrontato i risultati ottenuti.

Per quanto riguarda l'analisi qualitativa siamo riusciti a recuperare una serie di articoli cartacei che ci sono stati gentilmente offerti dalla Regione Puglia per effettuare tale analisi. Analizzato ogni singolo articolo siamo riusciti a ricavare una serie di topic.

Tale lavoro ha richiesto poco più di un mese in quanto sono stati analizzati più di cento articoli.

Per quanto riguarda, invece, l'analisi *Topic Modeling* per prima cosa bisognava recuperare gli articoli da analizzare in formato digitale.

Con l'utilizzo del Web oggi giorno abbiamo accesso ad una enorme quantità di informazioni e questo ci permette di recuperare qualsiasi dato a noi utile in un breve

lasso di tempo.

Tramite il sito *Il Sole 24 Ore* abbiamo recuperato, attraverso una ricerca avanzata, una serie di articoli inerenti alla Politica di Coesione.

Una volta raccolti i dati abbiamo iniziato la nostra analisi utilizzando algoritmi e software che in poco tempo hanno estratto una serie di topic contenente la lista di parole che caratterizzano maggiormente gli articoli raccolti.

Il software da noi utilizzato è il *Software Mallet* che ci ha permesso di ricavare una serie di Topic contenente l'elenco di parole che li caratterizzano.

Una volta concluse le due analisi ci siamo preoccupati di analizzare e confrontare i risultati ottenuti.

Come prima cosa possiamo subito dire che da un punto di vista più generale entrambi i metodi sono di gran lunga efficaci e indispensabili per effettuare l'analisi di un qualsiasi tipo di testo, ma, nessuno dei due è estremamente perfetto. Entrambi presentano lati positivi o negativi.

ANALISI QUALITATIVA	
ASPETTI POSITIVI	ASPETTI NEGATIVI
Meno strutturata	Maggiore concentrazione
Descrizione completa e dettagliata	Maggior tempo

Tab.9: tabella contenente gli aspetti positivi e negativi dell'analisi qualitativa dei dati.
Fonte: Ricostruzione personale.

L'*Analisi Qualitativa*, è meno strutturata rispetto all'*analisi Topic Modeling*, per questo permette di individuare una serie di sfumature di un determinato argomento che non potrebbero essere colte diversamente.

Infatti è stato molto più semplice associare un titolo e una breve descrizione ad ogni topic emerso da tale analisi in quanto analizzando l'intero articolo e non singole parole, si riesce a capire a pieno l'argomento trattato e le opinioni espresse a riguardo. L'aspetto negativo di tale analisi è che richiede molta concentrazione e molto tempo soprattutto se si parla di una quantità di articoli molto elevata.

ANALISI TOPIC MODELING	
ASPETTI POSITIVI	ASPETTI NEGATIVI
Ottimo strumento per analisi statistiche	Piena conoscenza dell'argomento in analisi
Ottimi risultati anche con testi poco puliti	Difficoltà interpretativa
Tempi molto veloci	Capire il numero giusto di topic da creare
Software facilmente utilizzabili	/
Analizza numeri ingenti di dati	/

Tab.10: tabella contenente gli aspetti positivi e negativi dell'analisi Topic Modeling dei dati.
Fonte: Ricostruzione personale.

L'analisi *Topic Modeling*, possiamo dire che sulla base dei risultati ottenuti, esso risulta un valido strumento per l'analisi statistica di dati testuali, nonostante alcuni aspetti richiedono maggiori ricerche e approfondimenti. Infatti, interpretare la lista delle parole ottenute non è molto semplice soprattutto se non si ha piena conoscenza dell'argomento che si sta trattando.

Alle volte, per associare un titolo ad un determinato topic si è dovuto ricercare e leggere l'articolo che aveva maggiore incidenza in quel topic così da essere certi di aver associato un titolo che descrivesse al meglio il topic. Un altro problema è cercare di capire il numero giusto di Topic da creare.

Questo è un fattore rilevante in quanto la bontà dei risultati dipende dall'individuazione di un numero adeguato di argomenti. Infatti, inizialmente abbiamo effettuato una analisi a 10 topic, una volta analizzato e confrontato i risultati ottenuti con quelli dell'analisi qualitativa ci è sembrato opportuno aumentare il numero di topic, in quanto, erano pochi i topic che risultavano simili tra le due analisi.

Così facendo abbiamo ricavato 15 topic e dal confronto con l'analisi qualitativa siamo riusciti a qualche somiglianza in più tra i vari topic.

I punti a favore di tale analisi, però, sono molteplici, ha la capacità di ottenere risultati apprezzabili anche senza un'onerosa, e spesso poco automatizzata, pulitura del

testo, ci permette di analizzare quantità ingenti di dati in tempi veramente brevi e con risultati incredibili.

In conclusione possiamo dire che i risultati ottenuti da entrambi i metodi sono assolutamente positivi, in quanto, in entrambi i casi siamo riusciti a ricavare i temi maggiormente affrontati e, nel caso dell'analisi qualitativa, le opinioni dei vari attori in merito ad essi.

L'analisi testuale è uno strumento molto potente e grazie all'espansione di internet e alla conseguente crescita delle informazioni disponibili in rete verrà reso sempre più importante lo studio di tecniche per l'elaborazione e la catalogazione di grandi quantità di dati testali.

Speriamo e ci auguriamo quindi, nel perfezionamento di queste tecniche tramite il miglioramento magari di software già esistenti, come Mallet, o alla creazione di altri, per far sì che questi studi vengano svolti in maniera più dettagliata, semplice e veloce.

BIBLIOGRAFIA

_AIELLO, Francesco, PUPO, Valeria, 2009. The Impact of European Cohesion Policy. A Study on Italian Regions. In *Rivista Italiana degli Economisti*, Vol. 3, N. 12, 2009. Pp. 421-454.

_BAUER, W., Martin, 2000. Classical Content Analysis: a Review. In BAUER, W. M. and GASKELL, G., (Eds). *Qualitative researching with text, image and sound*. Londra, Sage Publication. Pp. 131-151.

_BERELSON Bernard, 1952. *Content analysis in communication research*. New York, The Free Press.

_BEUGELSDIJK, Maaïke, and EIJJFINGER, C. W., Sylvester, 2005. The Effectiveness of Structural Policy in the European Union: An Empirical Analysis for the EU-15 in 1995-2001. In *Journal of Common Market Studies*, Vol. 43, No. 1, Pp. 37-51, (March 2005).

_BLEI, David M, Y. NG, Andrew, and JORDAN, Michael, I., 2003. Latent dirichlet allocation. In *The Journal of machine Learning research*, N. 3, Pp. 993–1022

_BLEI, David, M., 2012. Probabilistic topic models. In *Communications of the ACM*, 55(4), Pp. 77–84.

_BOLDRIN, Michele, CANOVA, Fabio, 2001. Inequality and Convergence in Europe's Regions: Reconsidering European Regional Policies. In *Economic Policy* N. 32, Pp. 53-75.

_CEE (Comunità Economica Europea), 1957. *Trattato che istituisce la Comunità economica europea (CEE)*. Roma, (25/3/1957).

_COMMISSIONE DELLE COMUNITÀ EUROPEE, 2008. *Libro verde sulla coesione territoriale. Fare della diversità territoriale un punto di forza*. Comunicazione della commissione al consiglio, al Parlamento Europeo, al Comitato Delle Regioni e al Comitato Economico e Sociale Europeo. Bruxelles, (6.10.2008), <http://ec.europa.eu/regional_policy/archive/consultation/terco/paper_terco_it.pdf>, (Ultimo Accesso, 12-06-2018).

_CRESPO CUARESMA, Jesus, **RITZBERGER-GRÜNWARD**, Doris and **SILGONER** Maria Antoinette, 2008. Growth, Convergence and EU Membership. In *Journal Applied Economics* Volume 40, Issue 5, Pp. 643-656.

_DULLI, Susi, **POLPETTINI**, Paola, **TROTTA**, Massimiliano, (a cura di), 2004. *Text mining: teoria e applicazioni*. Milano, Franco Angeli Editore.

_HOLSTI, Ole, 1968. Content Analysis. In **LINDZEY**, G. and **ARONSON**, E. (Eds), *The Handbook of Social Psychology*, Vol 2. Hoboken, New Jersey, John Wiley & Sons Inc.

_JARDINE, Nick, **VAN RIJSBERGEN**, Cornelis Joost, 1971. The use of hierarchical clustering in information retrieval. In *Information Storage and Retrieval*, vol. 7, Pp. 217–240.

_KOLLER Daphne, **FRIEDMAN** Nir, 2009. *Probabilistic Graphical Models. Principles and Techniques*. Cambridge, Massachusetts, USA, The MIT Press.

_KRIPPENDORFF, Klaus. 2004. Reliability in content analysis: some common misconceptions and recommendations. In *Human Communication Research*, Vol. 30(3), Pp. 411-433.

_LASSWELL, Harold Dwight, 1949. The Language of Power, In **LASSWELL**, H.D. e **LEITES**, Nathan, (Eds.), *Language of Politics; Studies in Quantitative Semantics*. New York, G.W. Stewart Publisher Inc. Pp-3-19.

_LOSITO, G., 1996. *L'analisi del contenuto nella ricerca sociale*. Milano, Franco Angeli Editore.

_LUHN, Hans Peter, 1959. Auto-encoding of documents for information retrieval systems. In **BOAZ**, Martha, (Editor), *Modern Trends in Documentation*. London, Pergamon Press. Pp. 45–58.

_MARON, Melvin Earl, **KUHNS**, John, L., 1960. On relevance, probabilistic indexing and information retrieval. In *Journal of the ACM*, vol. 7, Pp. 216–244.

_MORCELLINI, Mario e **FATELLI**, Giovanbattista, 1999. *Le Scienze della Comunicazione. Modelli e percorsi disciplinari*. Roma, Carocci Editore.

_SALTON, Gerard., 1964. A document retrieval system for man-machine interaction. In *Proceedings of the 19th Annual International ACM National Conference*. New York, ACM (Association for Computing Machinery) Press. Pp L2.3-1–L2.3-20, <http://delivery.acm.org/10.1145/810000/808923/pl2-3salton.pdf?ip=137.204.41.83&id=808923&acc=ACTIVE%20SERVICE&key=296E2ED678667973%2E274F812FD871B748%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1539603035_89e0217b422059393a83dfd6cb541fd6>, (Ultimo accesso: 24-06-2018).

_ŠPIDLA, Vladimír, 2008. 1989-93: Dai progetti ai programmi. In *Inforegio Panorama*. La politica di coesione dell'UE 1988-2008: investire nel futuro dell'Europa, n. 26, (giugno 2008), Pp. 8-13.<http://ec.europa.eu/regional_policy/sources/docgener/panorama/pdf/mag26/mag26_it.pdf>, (Ultimo accesso: 15-06-2018).

_VAROTTO, Mauro, 2010. *Le opportunità dell'Unione Europea per gli enti locali*. Milano, Ipsoa Edizioni.

_WIMMER, Roger and **DOMINICK**, Joseph, 1997. *Mass media research. An introduction*. Belmont (CA), Wadsworth Publishing Company.

SITOGRAFIA

_EUROPEAN COMMISSION.EU_REGIONAL POLICY INFOREGIO. *Financial Engineering Background: Jaspers-Jeremie-Jessica-Jasmine*, <http://ec.europa.eu/regional_policy/archive/funds/2007/jjj/index_en.htm>, (Ultimo accesso 23-06-2018).

MALLET.CS MACHine Learning for LanguagE Toolkit, <<http://mallet.cs.umass.edu/>>, (Ultimo accesso 12-06-2018).

TRECCANI.IT Enciclopedia della Matematica (2013), *Inferenza Statistica Bayesiana*, (voce in), <http://www.treccani.it/enciclopedia/inferenza-statistica-bayesiana_%28Enciclopedia-della-Matematica%29/>, (ultimo accesso: 12-11-2018).