



UNIVERSITÀ DI PISA

**DIPARTIMENTO DI INGEGNERIA DELL'ENERGIA DEI SISTEMI
DEL TERRITORIO E DELLE COSTRUZIONI**

**RELAZIONE PER IL CONSEGUIMENTO DELLA
LAUREA MAGISTRALE IN INGEGNERIA GESTIONALE**

***Metodologia per l'individuazione di PMI innovative in una
specifica area geografica attraverso tecniche di Text Mining***

SINTESI

RELATORI

Prof. Ing. Gualtiero Fantoni

*Dipartimento di Ingegneria Civile e Industriale,
Università di Pisa*

Dott. Nicola Melluso

*Dipartimento di Ingegneria dell'Energia dei Sistemi del Territorio E delle Costruzioni
Università di Pisa*

IL CANDIDATO

Gabriele Caruso

g.caruso12@studenti.unipi.it

Sessione di Laurea Magistrale del 18/06/2020
Anno accademico 2019/2020
Consultazione non consentita

Abstract

Il numero delle *PMI innovative* in Italia nell'ultimo anno ha registrato una sostanziale crescita. Il tessuto imprenditoriale italiano è destinato ad essere sempre più caratterizzato da questa tipologia di impresa. Una *PMI Innovativa* è uno status giuridico per l'impresa introdotto recentemente nell'ordinamento italiano a cui è dedicato uno specifico regime di agevolazioni che rappresenta un'opportunità strategica per l'imprenditore 4.0. Il numero di imprese attualmente certificate non rappresenta il numero di aziende *potenzialmente* innovative. Lo scopo di questa tesi è individuare quali imprese non certificate sono già in possesso dei requisiti necessari o che possono raggiungerne la conformità con minimi sforzi.

Nel presente lavoro è stata sviluppata una metodologia in grado di individuare quali imprese all'interno di un'area geografica siano potenzialmente *PMI innovative*. Il modello sviluppato per lo specifico obiettivo astrae il problema in tre dimensioni principali: settori merceologici, attività dell'impresa e struttura economico-patrimoniale. Tali dimensioni sono costruite attraverso l'utilizzo di tecniche statistiche e di text mining applicate su due campioni di riferimento: campione delle *PMI innovative certificate* e campione di *PMI generiche* relative ad una specifica area geografica. L'output della metodologia sarà un indice di sintesi riferito alla specifica PMI e una matrice in grado di guidare alla ricerca delle PMI innovative, *nascoste*, nell'area geografica di riferimento. Ciò potrà abilitare una comunicazione più efficace, sia a livello nazionale che locale, per il raggiungimento della policy e, a più alto livello, potrà avere un riscontro in tema business, per individuare tra i fornitori, competitor, potenziali entranti e clienti quali di essi offrano un maggior livello di innovazione abilitando così l'individuazione di *prospect* e/o benchmark.

In the last year the number of innovative SME in Italy has registered a significant growth. The Italian entrepreneurial net is doomed to be majorly characterised by this kind of enterprise. An innovative SME has recently been categorised as a legal entity in the Italian law, to which it has been dedicated specific benefit that could represent strategic opportunities for the entrepreneur 4.0. The number of company currently certificated does not represent the number of potentially innovative SME. This thesis aims at identifying which not certificated enterprises already possess some necessary requirements or can become compliant to the requirements effortlessly.

This work deals with a method that allows within a specific geographic area to identify which SMEs are potentially innovative. For this purpose, the method distinguishes three main dimensions: product sectors, enterprise activities and financial performances. These dimensions are obtained through the employment of statistical techniques and text-mining tools applied to two referential samples: one related to certified innovative SME and one of generic SME in a specific geographic area. The output of this method will be an index referring to a specific SME and a matrix that conducts to the research of "hidden" innovative SME in the graphic area analysed. this would lead to a more efficient communication both on national and local level in order to align with the policies. Furthermore, it could also have business consequences on the choice among suppliers, competitors, potential new entries and clients, those who can offer a major level of innovation so as to enable the identification of prospects and/or benchmarks.

1. Introduzione

1.1. Contesto

Lo status di “PMI innovativa” è previsto dall’ordinamento italiano da circa 5 anni. Nel 2019 sono poco meno di mille le aziende iscritte alla relativa sezione del *Registro delle Imprese*. A metà del 2020 sono 1466 registrando una crescita del 46% solo nell’ultimo anno. A questa nuova tipologia di azienda è dedicato uno speciale regime di agevolazioni, il quale non è sempre adeguatamente conosciuto, ma che, insieme ad altre esternalità positive associate al titolo di PMI innovativa, rappresenterebbe una grande opportunità di crescita sia dal punto di vista culturale che organizzativo per l’imprenditore 4.0. Insieme a fattori di contesto come le attuali dinamiche tecnologiche, lo scenario normativo e quello di mercato abilitano le condizioni di esistenza della tesi in oggetto, che fonda le basi sulla seguente ipotesi: oltre alle imprese già iscritte, esistono numerose PMI già in possesso dei requisiti per l’iscrizione al Registro delle PMI Innovative, o che possano raggiungerli con un’ottima qualità in termini di *costo-beneficio*.

1.2. Problema

Gli elementi che stanno rallentando la diffusione dell’attribuzione del titolo *PMI Innovativa* si possono identificare, da una parte, dalla scarsa “cognizione di sé” come innovativa da parte delle piccole medie imprese italiane, e, dall’altra, anche dalla scarsa capacità dal punto di vista comunicativo degli incentivi relativi, in particolare rivolti ai potenziali destinatari¹. Conoscere le PMI non certificate con le carte in regola per poterlo fare potrebbe abilitare una comunicazione più efficace alle parti interessate per il raggiungimento della policy, soprattutto perché si riuscirebbe a segmentare il mercato tra PMI innovative *certificate*, PMI *non innovative* e PMI innovative *certificabili*.

1.3. Obiettivo e Ambito di applicazione

L’obiettivo del lavoro di tesi è comprendere, all’interno di un’area geografica, quali siano le PMI generiche che hanno la potenzialità di diventare *PMI innovative* attraverso un metodo strutturato che offra una valutazione data driven sintetizzabile o in un indice finale o in una matrice che possa fungere da bussola per la ricerca delle PMI innovative non ancora certificate. La tesi in oggetto trova il suo campo di applicazione nelle PMI con sede legale in Toscana², in cui attualmente risiedono solo 74 PMI innovative certificate. La generalità della metodologia sviluppata consente la replicabilità ad altre aree geografiche o all’intero contesto nazionale, con il vincolo delle risorse computazionali a propria disposizione³. La metodologia proposta è fondata su modelli decisionali e di problem solving tipici dell’Ingegneria Gestionale supportati da tecniche statistiche tradizionali e di text mining.

¹ Alessandro Scutti, 2020 “PMI innovativa: uno status giuridico per l’impresa che guarda al futuro”

² Campione composto da 77587 PMI provenienti dal Database aziendale AIDA.

³ Soprattutto per quanto riguarda le operazioni di text mining

2. Metodologia e Applicazione

Per questioni di sintesi, la metodologia e il caso applicativo verranno illustrati in parallelo.

Il modello per l'estrazione delle misure è costruito attraverso tre livelli di astrazione⁴. Il primo livello consiste nel dividere lo spazio delle imprese in due archetipi⁵: PMI innovative certificate e PMI generiche (*senza certificazione*). Secondo l'ipotesi su cui si basa il lavoro di tesi, tra le PMI non certificate sono presenti degli *individui* che hanno caratteristiche per essere considerati PMI innovative. Nel secondo livello si riduce la complessità del problema individuando tre componenti principali, dimensioni sulle quali si desidera caratterizzare gli archetipi individuati su cui è possibile individuare delle metriche. Le dimensioni in oggetto sono: *settori merceologici, attività dell'impresa e struttura economico-patrimoniale*. Il terzo livello di astrazione consiste nell'applicazione, per analogia, del principio dell'analisi differenziale utilizzata nel controllo decisionale⁶: attraverso il confronto tra i due archetipi di impresa, per ogni dimensione definita, si isolano le caratteristiche peculiari e differenziali delle PMI innovative rispetto alle PMI generiche, ottenendo delle *metriche relative*. Si arriva così a costruire il modello, rappresentato qualitativamente in *Figura 1*, che culmina con la costruzione di uno spazio di misura⁷.

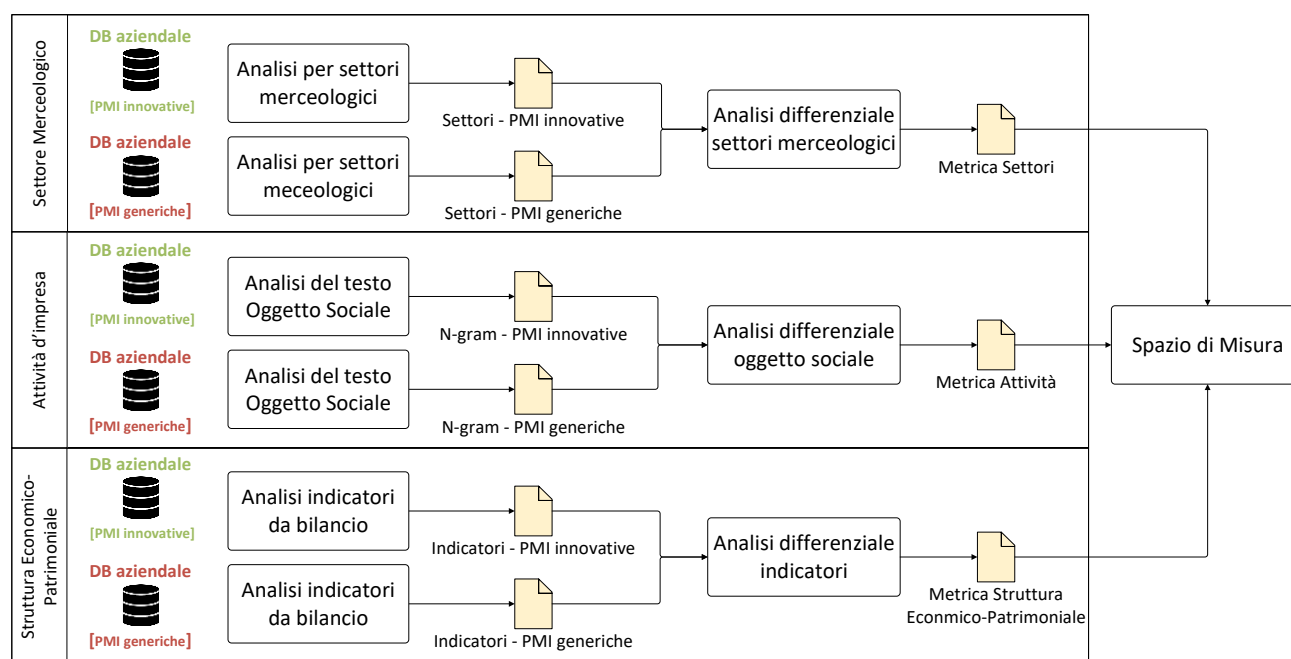


Figura 1- Mappa della metodologia

Come possibile notare dalla figura, per ogni dimensione definita, viene seguito lo stesso processo che permette di definire la metrica relativa. Nel caso applicativo, il campione delle PMI innovative è formato dalle 1466 aziende certificate allo status e il campione delle PMI generiche è composto da 77587 PMI toscane.

⁴ Tecnica di modellizzazione che consiste nell'abbassare il livello di dettaglio in modo da considerare solo elementi di interesse per i fini del modello

⁵ Archetipi sono aziende fittizie che, come le *personas* nella classica strategia di marketing, permette ai decisori di avvicinarsi alle caratteristiche della controparte reale attraverso uno sforzo di descrizione delle stesse, anche molto dettagliata. J Pruitt, J Grudin, 2003 "Personas: Practice and Theory"

⁶ Le applicazioni più famose di questa tecnica sono nelle decisioni di make or buy e dell'eliminazione di una linea di prodotto. M Maher, EB Deakin, 1994 "Cost accounting"

⁷ Donald L. Cohn, 1980 "Measure Theory" Boston, Birkhäuser, ISBN 0-8493-7157-0

2.1. Settori Merceologici

Concettualmente la dimensione *Settori Merceologici* vuole trasferire al modello l'informazione: se una generica PMI opera in un settore dove sono presenti un numero consistente di PMI innovative avrà più potenzialità di esserlo rispetto ad una PMI che opera in un settore scarso di questa tipologia di imprese. Tutto questo, come specificato sopra, in un'ottica differenziale. Per l'analisi dei settori vengono utilizzate le *divisioni* (livello di dettaglio: seconda cifra) della classificazione merceologica ATECO:2007.

I passaggi operativi per la costruzione della metrica, relativi a questa dimensione, sono:

1. Analisi della percentuale di aziende appartenenti al campione delle PMI innovative ($\theta_{PMI\ inn}$) e al campione delle PMI generiche ($\theta_{PMI\ i}$) per le divisioni ATECO:2007;
2. Individuazione, per ogni settore, di un peso equivalente alla differenza della percentuale di PMI innovative appartenenti alla divisione e la percentuale di PMI generiche appartenenti alla stessa (Δ_{s_i});
3. Individuazione della misura con cui un'azienda si posiziona rispetto a questa dimensione; questa è esattamente, il peso associato alla sua divisione di appartenenza ($\Delta_{s_j} = \Delta_{s_i}$):

$$\Delta_{s_j} = \theta_{PMI\ inn} - \theta_{PMI\ i} \quad \forall j \in S_i$$

$$i = 1, \dots, n$$

$$n = \text{numero settori merceologici}$$

$$j = 1, \dots, m_i$$

$$m_i = \text{numero di PMI generiche all'interno della divisione } i - \text{esima}$$

Nella Tabella 1 sono mostrati i pesi relativi alle prime 5 *divisioni* nel caso delle PMI toscane.

Settore Merceologico (<i>divisione</i>)	PMI innovative	PMI toscane	Δ_{s_j}
produzione di software, consulenza informatica e attività connesse	30.57%	1.56%	29.01%
ricerca scientifica e sviluppo	12.42%	0.33%	12.09%
fabbricazione di computer e prodotti di elettronica e ottica; apparecchi elettromedicali ed elettronici [...]	6.49%	0.42%	6.07%
attività di direzione aziendale e di consulenza gestionale	5.52%	2.34%	3.18%
attività degli studi di architettura e d'ingegneria; [...]	3.86%	1.14%	2.72%

Tabella 1- Primi cinque settori per il peso associato

Le PMI innovative italiane non sono presenti in 25 delle 86 divisioni ATECO:2007 in cui le PMI Toscane si distribuiscono. In *Figura 2* è mostrata la distribuzione standardizzata delle aziende per la misura in oggetto.

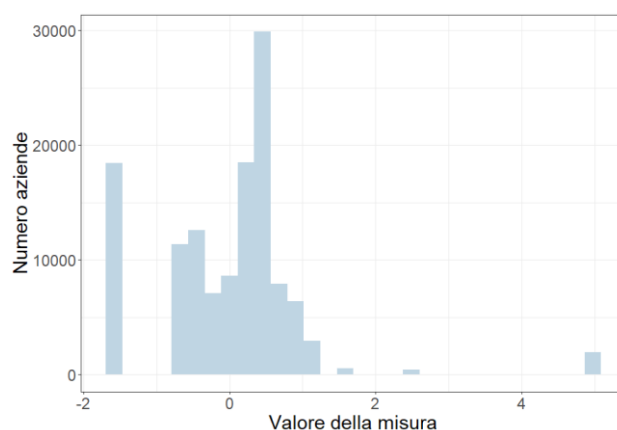


Figura 2- Distribuzione delle PMI Toscane per Δ_{s_j} standardizzato

La dimensione è quindi caratterizzata da valori discreti e crescenti con il livello di potenzialità corrispondente. Una misura di quanto il settore merceologico sia una dimensione in cui trovare aspetti differenziali tra gli archetipi in oggetto è la divergenza tra la distribuzione delle PMI all'interno dei settori per i diversi campioni.

2.2. Attività dell'impresa

L'Attività dell'impresa è la dimensione volta ad esplorare le modalità con cui una specifica PMI svolge le proprie attività di business. Si ipotizza che l'oggetto sociale⁸ sia una delle possibili fonti per questa tipologia di informazioni. Anche per questa dimensione l'interesse è quello di catturare le modalità differenziali con cui una PMI innovativa svolge il proprio business rispetto alle imprese generiche, ovvero, individuare entità testuali che siano peculiari e differenziali per gli oggetti sociali delle PMI innovative.

Infatti, trattandosi di una fonte disponibile sottoforma testuale, l'estrazione della conoscenza avviene attraverso un processo di Text Mining⁹. Operativamente, per ognuno dei due campioni di aziende relativi agli archetipi, per abilitare la successiva analisi differenziale, si eseguono le attività in Tabella 2.

Nome dell'attività	Descrizione dell'attività	Descrizione output
Tokenizzazione	L'oggetto sociale è un documento. I documenti sono informazioni non strutturate, questi devono essere divisi in unità linguistiche (token). Le parole vengono separate l'una dall'altra dallo "spazio bianco", ma la regola dello spazio bianco non è sempre sufficiente per avere una tokenizzazione adeguata.	"I" "azienda" "offre" "." "prodotti" "ad" "elevato" "contenuto" "tecnologico"
Text cleaning e eliminazione stopwords	I token individuati dall'attività precedente vengono depurati da entità con valore informativo nullo: <i>stopwords</i> , lista di parole presenti in ogni documento, e segni di punteggiatura, simboli, numeri ecc...	"il" "lo" "essere" "un" "_" "."
Estrazione di entità testuale (n-grams)	Un n-gram è una sequenza consecutiva di n parole. Il metodo adottato consiste nell'estrazione di tutti i possibili n-grams (con n=2 e n=3) e caratterizzarli in base alla loro frequenza all'interno degli oggetti sociali del campione oggetto di studio	"alto contenuto tecnologico" "servizi innovativi" "nuove tecnologie" "servizi informatici"
Rimozione di entità testuali non significative	Sono rimosse le entità testuali: a) che si verificano in meno dell'1% dei documenti; b) che sono terminologia tipica del documento sotto analisi	"attività" "oggetto sociale" "potrà compiere"

Tabella 2- Attività sugli Oggetti Sociali per ogni campione relativo agli Archetipi d'impresa

L'output di questo processo è individuato dalle entità testuali e la loro frequenza all'interno del campione delle PMI innovative e all'interno del campione delle PMI toscane. A questo punto, il principio dell'analisi differenziale si applica confrontando la frequenza delle entità testuali, all'interno degli oggetti sociali, in comune ad entrambi i campioni. Le attività eseguite sono descritte in Tabella 3:

⁸ L'oggetto sociale è un testo che descrive l'attività primaria economica che un'azienda si prefissa di perseguire durante la sua vita. È obbligatoria secondo il Codice Civile italiano e deve essere emanata nel momento dell'atto costitutivo.

⁹ Julia Silge and David Robinson, 2020- "Text Mining with R- A Tidy Approach"

Nome dell'attività	Descrizione dell'attività	Descrizione output
Intersezione e filtraggio per frequenza relativa	Viene effettuata l'intersezione tra le entità testuali in comune ad entrambi i campioni di PMI. Tra queste sono selezionate solo quelle caratterizzate da una frequenza maggiore nel campione delle PMI innovative rispetto a quelle generiche. D'ora in poi, questa differenza di frequenze verrà chiamata <i>frequenza relativa</i> (v_{rel}).	Lista delle entità testuali più frequenti negli oggetti sociali delle PMI innovative
Individuazione delle keyword	Le entità testuali vengono ulteriormente filtrate: a) attraverso analisi di concentrazione su indicatori quantitativi basati sulla frequenza. b) Eliminando duplicati di bi-gram e tri-gram	<i>Keyword (816)</i> : elementi testuali differenziali definitivi caratterizzati dalla propria <i>frequenza relativa</i> (v_{rel})
Individuazione delle misure	Ogni keyword individuata viene misurata rispetto ad una specifica azienda in base alla frequenza della stessa nel relativo oggetto sociale (<i>term-frequency</i> , <i>tf</i>) pesata per l'occorrenza negli oggetti sociali in generale (<i>inverse document frequency</i>) ¹⁰	Matrice che ha come righe le PMI toscane e come colonne le <i>keyword</i> . Nell'incrocio c'è il valore $tf*idf$ che assume la keyword all'interno dell'oggetto sociale della specifica azienda

Tabella 3-Attività per individuare metriche di differenziazione per la dimensione in oggetto

Nella *Figura 3* sono mostrate, a sinistra, le entità testuali peculiari e differenziali (*keywords*) per le PMI innovative con la relativa misura della frequenza relativa e della frequenza assoluta negli oggetti sociali delle PMI innovative; a destra è mostrata la distribuzione dell'occorrenza delle *keywords* tra gli oggetti sociali delle PMI toscane: circa il 30% delle aziende all'interno del campione possiede una sola *keyword* almeno una volta all'interno dell'oggetto sociale e, circa il 50% di esse, ne contiene solo una con una sola occorrenza.

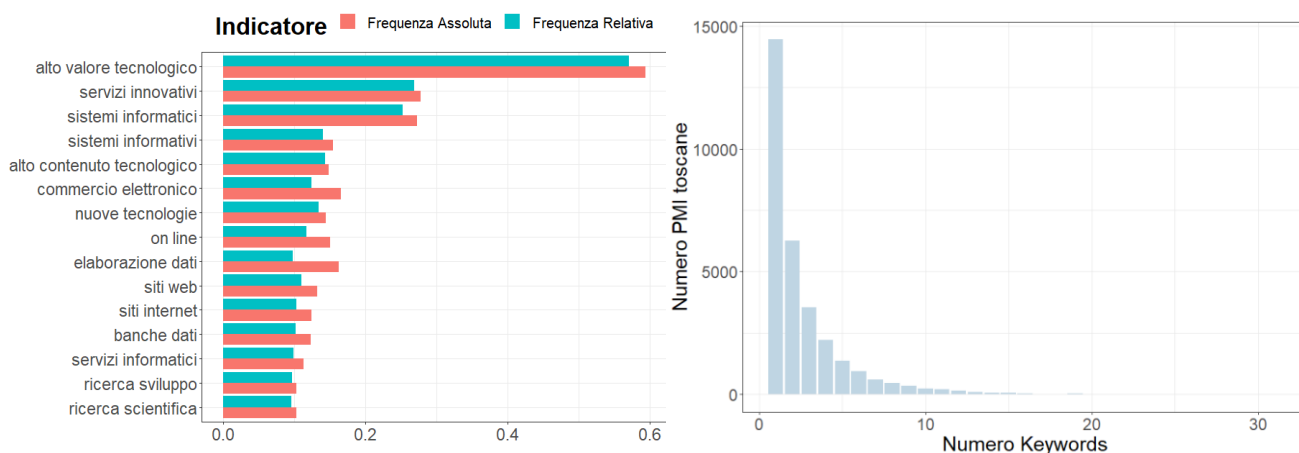


Figura 3- Risultati del processo di Text Mining

Concludendo, La misura con cui un'azienda si posiziona rispetto a questa dimensione sarà composta dalla sommatoria dei valori tf_idf delle *keywords* all'interno dell'oggetto sociale di quella azienda; ogni valore della sommatoria è pesato per il valore della frequenza relativa riferito alla keyword specifica, come mostrato nella seguente equazione.

¹⁰ Benoit, Kenneth, and Paul Nulty. 2016. *quanteda: Quantitative Analysis of Textual Data* & Julia Silge and David Robinson, 2020- "Text Mining with R- A Tidy Approach"

$$\Delta_{A_j} = \sum_{i=1}^n tf_idf_{ji} \cdot v_{rel_i}$$

$i = 1, \dots, n$

$n = \text{numero keywords individuate}$

$j = 1, \dots, m$

$m = \text{numero PMI generiche all'interno del campione}$

Come già si poteva intuire dalla parte destra della Figura, la distribuzione della misura per il campione delle PMI toscane è molto concentrata, lo testimonia un indice di Gini di 0.89.

2.3. Struttura economico-patrimoniale

La terza ed ultima dimensione ha il fine di catturare l'informazione riguardante gli aspetti peculiari e differenziali delle PMI innovative per la struttura economica-patrimoniale declinata sul tema dell'innovazione. Sono due gli indicatori individuati per perseguire l'obiettivo, questi sono stati definiti in maniera deduttiva dai requisiti richiesti per la certificazione allo status di PMI innovativa.

Requisiti status PMI innovative	Indicatore	Ipotesi per l'indicatore
Volume di spesa in Ricerca, Sviluppo e Innovazione in misura uguale o superiore al 3% della maggiore entità fra costo e valore totale della produzione;	Totale immobilizzazioni immateriali sul totale dell'attivo	Un'azienda con spese in ricerca e sviluppo impiega fonti nella voce "costi di sviluppo". ¹¹
Titolarietà, anche quali depositarie o licenziatarie, di almeno una privativa industriale [...]		All'interno della voce dello stato patrimoniale immobilizzazioni immateriali è presente la sotto-voce "diritti di brevetto industriale e diritti di utilizzazione delle opere dell'ingegno"
Impiego di dipendenti o collaboratori, in percentuale $\geq$ a 1/5 forza lavoro complessiva, di personale in possesso di titolo di dottorato di ricerca o che stia svolgendo un dottorato di ricerca. Oppure $\geq$ a 1/3 con laurea magistrale	Costo per salari e stipendi sul costo totale operativo	Il valore monetario dello stipendio di un dipendente è proporzionale all'education, senza considerare il fattore esperienza. ¹²

Tabella 4- Ipotesi per la definizione degli indicatori da bilancio

Individuati gli indicatori di interesse, è necessario indagare quali valori siano tipici per il campione delle PMI innovative. Affinché le ipotesi legate agli indicatori siano plausibili ci si attende che per le aziende innovative certificate il valore degli stessi siano tendenzialmente maggiori rispetto alle PMI generiche.

Operativamente si concretizza attraverso le fasi:

1. Studio della distribuzione del valore per i due indicatori nella popolazione delle PMI innovative e individuare, in base alla stessa, delle classi omogenee di imprese;
2. Per le classi definite, valutare la distribuzione dei valori per il campione delle PMI generiche toscane;
3. La metrica relativa per la dimensione in oggetto è costruita valutando la differente percentuale di popolazione all'interno delle classi definite tra le PMI innovative ($\mu_{PMI\ inn}$) e le PMI toscane (μ_{PMI}).

¹¹ L'informazione diretta per il requisito è contenuta nella nota integrativa per cui, attualmente, non si ha accesso. Potrebbe rappresentare un'importante fonte per sviluppi futuri. Un'altra informazione utile presente nel documento è il numero di dipendenti impiegati nell'esercizio.

¹² L'assunzione in questo caso è piuttosto forte

Questa differenza è mostrata nella *Figura 4*.

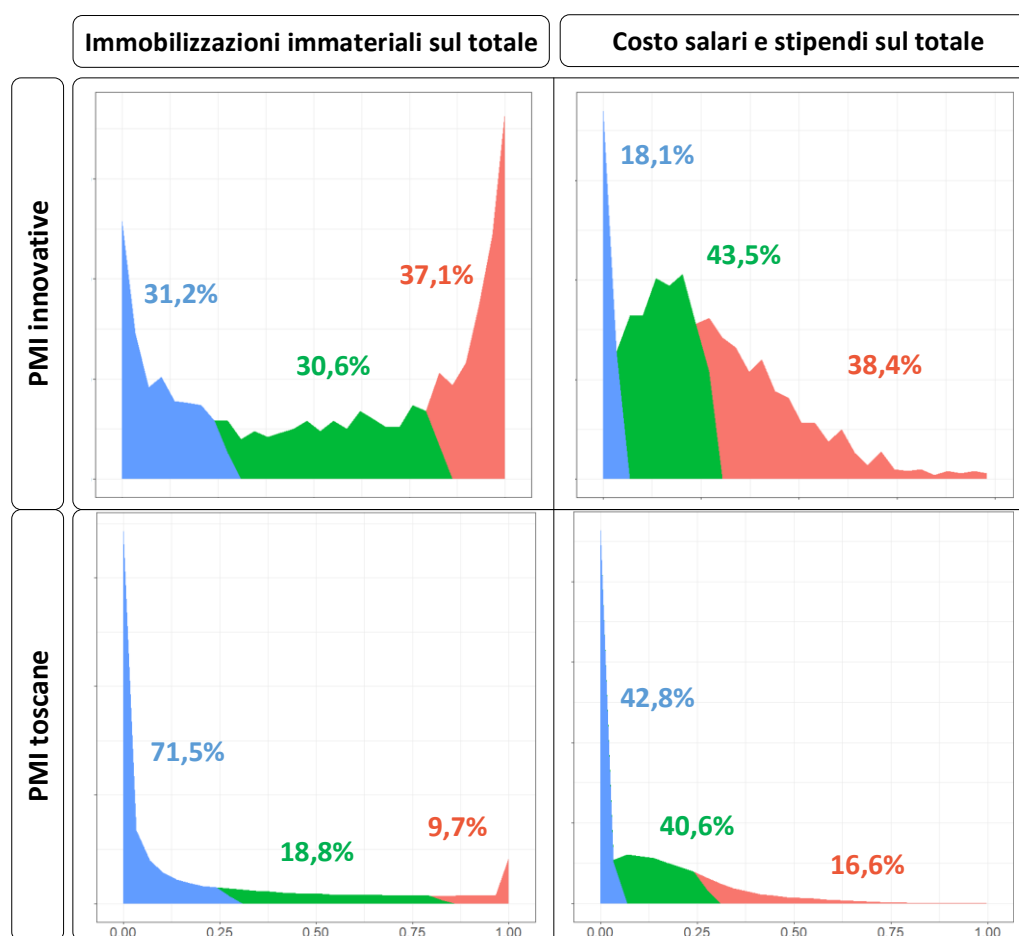


Figura 4- Distribuzione del valore degli indicatori per i due Archetipi d'impresa

È possibile notare dalla *Figura 4* come la percentuale di PMI innovative, per le classi in cui il valore dell'indicatore è più elevato, è maggiore rispetto al campione delle PMI toscane. La metrica differenziale è costruita con la stessa logica dei settori merceologici, assegnando alla classe un peso dato dalla sottrazione delle percentuali di PMI nei due diversi campioni, come mostrato in *Tabella 5*.

Classe	Immobilizzazioni immateriali sul totale			Costo salari e stipendi sul totale		
	PMI innovative	PMI toscane	Δ_{P_i}	PMI innovative	PMI toscane	Δ_{P_i}
A	37,1%	9,7%	27,3%	38,4%	16,6%	21,8%
B	30,6%	18,8%	11,8%	43,5%	40,6%	2,9%
C	31,2%	71,5%	-39,8%	18,1%	42,8%	-24,7%

Tabella 5- Metriche per la dimensione *Struttura economico-patrimoniale*

Infine, la misura con cui un'azienda si posiziona rispetto a questa dimensione sarà il peso associato alla classe di appartenenza del valore assunto dalla stessa per l'indicatore specifico ($\Delta_{P_j} = \Delta_{P_i}$):

$$\Delta_{P_j} = \mu_{PMI\ inn_i} - \mu_{PMI_i} \quad \forall j \in C_i$$

$i = 1, \dots, n$
 $j = 1, \dots, m_i$

n = numero classi definite per l'indicatore
 m_i = numero di PMI generiche all'interno della classe i – esima

2.4. Spazio di misura

Ogni metrica definita per la specifica dimensione entra in input per la costruzione di uno spazio tridimensionale di misura. Lo spazio di misura è sintetizzato in *Tabella 6*.

Dimensione	Misure	Fattori per metriche
Settori Merceologici	$\Delta_{S_j} = \theta_{PMI\ inn_i} - \theta_{PMI_i} \quad \forall j \in S_i$ $i = 1, \dots, n \quad n = \text{numero settori merceologici}$ $j = 1, \dots, m_i \quad m_i = \text{numero di PMI generiche all'interno della divisione } i - \text{esima}$	Percentuale delle PMI innovative per ogni settore merceologico ($\theta_{PMI\ inn_i}$)
		Percentuale delle PMI generiche per ogni settore merceologico (θ_{PMI_i})
Attività dell'impresa	$\Delta_{A_j} = \sum_{i=1}^n tf_idf_{ji} \cdot v_{rel_i}$ $i = 1, \dots, n \quad n = \text{numero keyword individuate}$ $j = 1, \dots, m \quad m = \text{numero PMI generiche all'interno del campione}$	Valore del tf-idf per ogni keyword all'interno dell'oggetto sociale della PMI generica i-esima (tf_idf_{ij})
		Frequenza relativa per ogni keyword i-esima identificata (v_{rel_i})
Struttura economico-patrimoniale	$\Delta_{P_j} = \mu_{PMI\ inn_i} - \mu_{PMI_i} \quad \forall j \in C_i$ $i = 1, \dots, n \quad n = \text{numero classi definite per l'indicatore}$ $j = 1, \dots, m_i \quad m_i = \text{numero di PMI generiche all'interno della classe } i - \text{esima}$	Percentuale delle PMI innovative nelle classi per ogni indicatore ($\mu_{PMI\ inn_i}$)
		Percentuale delle PMI generiche nelle classi per ogni indicatore (μ_{PMI_i})

Tabella 6- Sintesi dello spazio di misura

Date le misure effettuate per ogni PMI generica, l'interpretazione dei risultati può avvenire secondo due approcci: quantitativo e/o qualitativo. Con l'approccio quantitativo vi è un'aggregazione delle misure per ogni dimensione in un unico indice che esprima la potenzialità di un'impresa di essere classificata come PMI innovativa. Tipicamente raccogliere tutto in un unico indice può portare a delle distorsioni e, inoltre, può far perdere l'informazione del contributo che ogni dimensione apporta allo stesso. Alla luce di questo, con un approccio qualitativo si offre la possibilità di valutare la potenzialità di una impresa di essere PMI innovativa mantenendo la visione dei singoli contributi dimensionali. L'approccio quantitativo è costruito, invece, vedendo lo spazio di misura come sistema cartesiano a tre assi (dimensioni) in cui le metriche *relative* individuate per ogni asse sono l'unità con cui vengono misurate le generiche PMI. Le dimensioni sono costruite in modo tale che, maggiore è il valore associato ad un'azienda, più elevata è la potenzialità che essa sia innovativa; di conseguenza, l'indice finale può essere rappresentato, come si può notare nella formula sotto riportata, previa traslazione del sistema di riferimento, dalla *distanza euclidea* tra il punto dello spazio individuato dei valori assunti dalla PMI j-esima e il punto individuato dall'origine degli assi.

$$\Delta_j = \sqrt{\Delta_{S_j}^2 + \Delta_{A_j}^2 + \Delta_{P_j}^2}$$

Dove Δ_{P_j} è la risultante della media tra le misure dei due indicatori economico-patrimoniale.

Come già anticipato questo approccio ha diverse criticità, una delle quali, quella di considerare tutte le dimensioni con la stessa importanza. Un valore consistente per la dimensione *attività d'impresa*, secondo il modello costruito, deve essere una condizione necessaria per la valutazione della specifica impresa: l'indice Δ_j potrebbe assumere un alto valore solo per i contributi delle altre due dimensioni; per fare in modo che ciò non accada, l'approccio quantitativo viene applicato solo al 20% delle PMI toscane che, secondo l'analisi di

concentrazione, cattura il 95% della cumulata di Δ_{A_j} (indice di Gini 0.89). Inoltre, queste criticità offrono la spinta per la definizione del secondo approccio. L'approccio qualitativo ha l'obiettivo di evitare tutte le distorsioni che derivano dall'aggregazione delle diverse misure in un unico indice e rappresentare una bussola che guidi il decisore nel processo di indagine delle PMI *innovative*. La rappresentazione qualitativa dello spazio di misura consiste in una matrice, in *Figura 6*, in grado di mappare le tre dimensioni in un piano¹³.

		Settori «Tradizionali»			Settori «Innovativi»		
		Costo salari sul totale			Costo salari sul totale		
		C	B	A	C	B	A
Oggetto Sociale «Tradizionale»	Immobilitazioni immateriali sul totale	17892	17228	6192	1437	1859	824
	B	2999	5386	1943	271	512	201
	A	2546	1418	778	325	140	113
Oggetto Sociale «Innovativo»	Immobilitazioni immateriali sul totale	2666	2808	1021	1529	1167	844
	B	755	1002	349	407	450	308
	A	703	320	178	561	213	242

Figura 5- Matrice qualitativa rappresentante lo spazio di misura

Le metriche relative per l'attività dell'impresa sono suddivise considerando un oggetto sociale "innovativo" e un oggetto sociale "tradizionale"¹⁴. Lo stesso passaggio è effettuato per le metriche associate ai settori¹⁵. Per gli indicatori *economico-patrimoniali* le classi sono già state identificate. Nell'incrocio tra le categorie dimensionali definite è specificato, attraverso un cerchio, il numero di PMI che vive in quel determinato segmento¹⁶. Ogni segmento è valorizzato dalle metriche per ogni dimensione e, tendenzialmente, muovendosi lungo la diagonale, da sinistra verso destra e dall'alto verso il basso, aumenta la potenzialità del segmento, intesa come possibilità che all'interno dello stesso viva una *PMI innovativa nascosta*, codificato anche dal colore più scuro del cerchio. Si può notare come la tonalità scura non sia lineare con la direzione della diagonale; questo è dovuto proprio al fatto che il modello costruito non dà la stessa importanza alle tre dimensioni. A differenza della metodologia quantitativa, è possibile entrare nel merito di una singola PMI e valutarne la possibilità che questa possa essere *PMI innovativa* senza considerare necessariamente le dimensioni in toto. A titolo di esempio, un decisore potrebbe valutare una *PMI* che presenta una misura

¹³ La figura mostra l'applicazione della matrice per il caso studio della Toscana

¹⁴ Categorie emerse da distribuzione paretiana rispetto a Δ_{A_j} : il 20% delle PMI che catturano il 95% del valore cumulato Δ_{A_j} sono considerate aventi oggetto sociale "innovativo"

¹⁵ Settori "innovativi" sono quelli caratterizzati da una percentuale maggiore di PMI innovative rispetto alle generiche

¹⁶ Il diametro del cerchio è proporzionale con il numero delle PMI all'interno del segmento

dell'oggetto sociale elevata, ma che, attualmente, stia operando in un settore “*tradizionale*” o che non presenti indicatori economico patrimoniali di classe A. All'interno del singolo segmento sarà poi possibile applicare la metodologia quantitativa, la distanza euclidea, per effettuare delle classifiche locali.

Con riferimento al Ciclo di Deming (*PDCA*), un importante strumento di validazione consisterà nel monitoraggio dei segmenti nel quale vivono le PMI che via via si certificano come *innovative* (*Check*), per poi utilizzare questa informazione per aggiornare il modello in un'ottica di miglioramento continuo (*Act*).

3. Conclusioni e sviluppi futuri

Il lavoro di tesi propone una metodologia strutturata costruita attraverso un modello *top-down*, con riferimento alle caratteristiche peculiari delle *PMI innovative*, che individua per ogni azienda generica all'interno di una specifica area geografica la potenzialità che, nella stessa, si nasconde un ulteriore PMI innovativa. La metodologia si concretizza in un indice finale e/o in una matrice che propone una mappatura delle aziende rispetto alle dimensioni di valutazione. Il secondo output rappresenta uno strumento in grado di supportare un processo di *decision-making*, unendo ai numeri l'esperienza e la conoscenza del contesto dell'essere umano. Il modello, attualmente in stato prototipale, ha ampi margini di miglioramento:

- A livello tecnico-algoritmico è possibile sostituire lo spazio di misura con un algoritmo di classificazione che, in base alle misure effettuate, anche meno aggregate, potrebbe restituire un indice di probabilità che un'azienda sia una *PMI innovativa*, basandosi su *fitness* matematica, con conseguente aumento di automazione¹⁷. Con l'algoritmo, però, si abiliterebbero alcuni meccanismi *black-box* perdendo il controllo sui risultati presente nell'attuale metodologia. Solito *trade-off* tra automazione e presidio dell'output;
- A livello della qualità degli input è possibile sostituire l'oggetto sociale, che contiene informazioni statiche e di livello generale, con le informazioni presenti sui siti-web aziendali, più specifiche e dinamiche.

Su questo campo sono stati effettuati vari test dal sottoscritto. Il vantaggio è che le informazioni in input sono di maggior dettaglio e maggior valore informativo, ma ci sono anche diversi svantaggi: la copertura in Italia delle aziende con sito web è bassa e, oltre a questo, l'oggetto sociale ha il grande vantaggio di trovarsi in tabella, risparmiando l'impiego di risorse consistenti nell'attività di *information extraction* (*IE*)¹⁸.

- Sempre rispetto alla qualità degli input, per la dimensione riguardante gli indicatori di bilancio, le ipotesi di costruzione degli indicatori potranno evolversi considerando anche gli attuali lavori di ricerca su queste tematiche¹⁹.

Il modello applicato solo alle PMI toscane, oltre all'immediato potere di replicabilità in tutto il contesto nazionale, in futuro, con le dovute modifiche, potrà essere applicato per ricercare *tipologie* di imprese diverse dalle *PMI innovative*, sostituendo quest'ultimo archetipo dagli input con un campione contenente la *tipologia di azienda* di interesse: fornitori, clienti, competitor, ecc...

¹⁷ Per realizzare il classificatore, però, oltre alla conoscenza di quelle che sono già *PMI innovative* è necessaria l'informazione di quelle che certamente non lo sono, per costruire un adeguato *training set*

¹⁸ Senza considerare il fatto che l'attività di *Text Mining* vera e propria, nel caso dei siti web, sarebbe molto più dispendiosa a livello computazionale e di progettazione.

¹⁹ Carlo Alberto Magni, 2019 “*Investment Decisions and the Logic of Valuation*”