



UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA “GIORGIO FUÀ”

Corso di Laurea Magistrale in Data Science per l’Economia e le Imprese
Interclasse LM-56 e LM-91

ANALISI DEGLI INDICATORI DI
ALLERTA PER LA CRISI D’IMPRESA:
L’APPLICAZIONE DEL MODELLO CNDCEC

ALERT INDICATORS FOR BUSINESS CRISIS:
THE CNDCEC MODEL

Relatore: Prof.ssa
Maria Cristina Recchioni

Tesi di Laurea di:
Fabien Gripshi

Correlatore: Prof.
Giovanni Marco Borroni

Anno Accademico 2021 – 2022

Sommario

INTRODUZIONE	7
1 IL CONCETTO DI CRISI	9
1.1 LA DEFINIZIONE DI CRISI D'IMPRESA	9
1.2 EVOLUZIONE DELL'APPROCCIO METODOLOGICO PER LA PREVISIONE DELLA CRISI.....	11
2 IL MODELLO CNDCEC	14
2.1 DESCRIZIONE DEL MODELLO REALIZZATO DAL CNDCEC.....	14
2.2 APPLICAZIONE DEL MODELLO CNDCEC	23
2.3 RISULTATI OTTENUTI	37
2.4 OSSERVAZIONI E COMMENTI SUI RISULTATI	40
3 MACHINE LEARNING AI FINI DELLA PREVISIONE DELLA CRISI.....	51
3.1 CRITICITÀ NELLA PREVISIONE DELLA CRISI D'IMPRESA.....	51
3.2 ANALISI DISCRIMINANTE LINEARE.....	55
3.3 COMMENTO DEI RISULTATI OTTENUTI E CONCLUSIONI.....	58
CONCLUSIONI E SVILUPPI FUTURI.....	64

Elenco delle tabelle

TABELLA 2.1: AGGREGAZIONI DEI SETTORI SECONDO LE DEFINIZIONI ATECO 2007	19
TABELLA 2.2: SOGLIE SETTORIALI	21
TABELLA 2.3: DESCRIZIONI DELLE VARIABILI DELLO STATO PATRIMONIALE.....	27
TABELLA 2.4: DESCRIZIONI DELLE VARIABILI DEL CONTO ECONOMICO	28
TABELLA 2.5: RISULTATI QUERY DI RICERCA.....	29
TABELLA 2.6: SUDDIVISIONE IMPRESE CON PATRIMONIO NETTO NEGATIVO	31
TABELLA 2.7: CONTEGGIO SETTORIALE DELLE IMPRESE	34
TABELLA 2.8: VALORI NULLI PER INDICE.....	36
TABELLA 2.9: NUMERO DI IMPRESE CHE SUPERANO INDICATORI CON TASSO DI DEFAULT.....	37
TABELLA 2.10: MATRICE DI CONFUSIONE DEGLI INDICI SECONDARI DEL MODELLO CNDCEC.....	37
TABELLA 2.11: MISURE DI PERFORMANCE DEGLI INDICI SECONDARI DEL MODELLO CNDCEC	37
TABELLA 2.12: MISURE DI PERFORMANCE DI COMBINAZIONI SCELTE	40

TABELLA 2.13: STATISTICHE DESCRITTIVE DEGLI INDICATORI PER CLASSE	45
TABELLA 2.14: CONFUSION MATRIX E PERFORMANCE COMPLESSIVE DEL MODELLO CNDCEC	50
TABELLA 3.1: CONFRONTO TRA PERFORMANCE DEL MODELLO CNDCEC E MODELLO LDA	58
TABELLA 3.2: COEFFICIENTI DELLA COMBINAZIONE LINEARE DELLA LDA.....	59
TABELLA 3.3: PERFORMANCE DEL MODELLO A SEGUITO DELLA RFE	62
TABELLA 3.4: VARIABILI E PESI DEL MODELLO DERIVANTE DALLA LDA CON RFE	62

Elenco delle figure

FIGURA 2.1: DISTRIBUZIONE DEL NUMERO DI INDICI SUPERATI DA CIASCUNA IMPRESA	36
FIGURA 2.2: DISTRIBUZIONE DELL'INDICE DI LIQUIDITÀ 2017, PER CLASSE.....	44
FIGURA 2.3: DISTRIBUZIONE DELL'INDICE DI INDEBITAMENTO PREVIDENZIALE E TRIBUTARIO 2017, PER CLASSE	45

INTRODUZIONE

Il progetto di ricerca esposto in questa tesi affronta la tematica della crisi d'impresa. Questo argomento è contraddistinto da un'ampia rilevanza economica e sociale, oltre ad essere dibattito di grande attualità, per una serie di motivi che verranno elencati in questa breve introduzione. Con la pubblicazione in Gazzetta Ufficiale del Decreto Legislativo numero 14 del 12 gennaio 2019, in attuazione della legge 19 ottobre 2017 numero 155, il 14 febbraio 2019 è entrato in vigore il Codice della Crisi e dell'Insolvenza d'Impresa.^[1] Il nuovo Codice, applicato effettivamente solo a partire dal 1° settembre 2021 a causa dell'emergenza sanitaria Covid-19, si pone l'obiettivo di riformare in maniera unitaria ed organica la disciplina delle procedure concorsuali e della crisi dovuta al sovraindebitamento, semplificando l'impianto normativo e garantendo la certezza del diritto. Una delle novità più importanti della riforma è l'introduzione all'istituzione di una procedura di allerta e di composizione assistita della crisi, con il compito di creare uno strumento di supporto diretto ad analizzare le cause della difficoltà economica e finanziaria dell'impresa per eventualmente prevenire l'ingresso dell'impresa nello stato di crisi. La crescente disponibilità di dati relativi ai vari aspetti dell'attività dell'impresa, insieme al progresso tecnico in merito

all'analisi di questi dati, rende possibile la creazione di queste procedure di allerta, concretizzando modelli di previsione della crisi d'impresa. Queste motivazioni hanno condotto alla volontà di approfondire l'argomento, analizzando gli sviluppi più recenti e fornendo basi teoriche e pratiche per la comprensione della materia.

L'obiettivo principale sarà quello dell'applicazione del modello elaborato seguendo le disposizioni di legge, valutandone l'efficacia nella rilevazione della crisi d'impresa. Ai fini di una valutazione comprensiva, verrà eseguito un confronto con semplici modelli di *machine learning*.

Nel primo capitolo, "*Il concetto di crisi*", verrà spiegata la crisi nelle sue accezioni. Verrà poi presentata nel secondo capitolo, "*Il modello CNDCEC*", la più recente modellistica disponibile per la previsione e l'analisi della crisi d'impresa. Nel terzo capitolo, "*Machine Learning ai fini della previsione della crisi*", verrà introdotto un approccio personale mirato alle risoluzioni delle problematiche presenti nella previsione della crisi stessa. Infine, verranno espone opinioni e idee per un possibile sviluppo futuro dell'argomento, derivanti dai risultati delle analisi effettuate e dalle considerazioni sostenute in questo elaborato.

Capitolo 1

IL CONCETTO DI CRISI

Nel presente capitolo verrà descritto in dettaglio l'argomento di questo progetto di ricerca, in modo da definire il dominio applicativo in cui sarà condotta l'analisi desiderata. Verrà quindi definita la crisi d'impresa e come il concetto di crisi e della sua previsione si sia evoluto negli anni.

1.1 LA DEFINIZIONE DI CRISI D'IMPRESA

È necessario innanzitutto comprendere il concetto teorico di crisi d'impresa per spiegare l'attualità e la rilevanza sociale del problema posto in analisi. In ottica aziendalistica, la crisi si identifica come il venir meno delle condizioni di equilibrio economico e finanziario dell'impresa capaci di compromettere la prospettiva di continuità aziendale.^[2] All'art. 2 del Codice della Crisi e dell'Insolvenza d'Impresa (CCII), la crisi, viene definita "come lo stato di squilibrio economico-finanziario che rende probabile l'insolvenza del debitore, e che per le

imprese si manifesta come inadeguatezza dei flussi di cassa prospettici a far fronte regolarmente alle obbligazioni pianificate”.^[1] L'imprenditore ha il compito di monitorare la sussistenza di tali minacce e valutare eventuali rimedi; gli organi di controllo hanno il compito di verificare il costante monitoraggio da parte dell'imprenditore¹. Questo punto di attenzione richiamato dal Legislatore rappresenta una categoria più ampia della non sostenibilità dei debiti. La non sostenibilità dei debiti è certamente una minaccia alla continuità, ma vi possono essere minacce correlate ad altro tipo di eventi. Alcuni di questi sono intercettabili da un affidabile sistema di *risk management*, tra cui controversie giudiziarie che coinvolgono i vertici della società, oppure perdita improvvisa di clienti o fornitori fondamentali. Nella presente analisi si fa riferimento alla crisi d'impresa come un processo che può crearsi per svariati motivi di gestione aziendale che, aggravandosi, nell'atto finale può culminare in insolvenza.

¹ Ex art. 2086 del Codice Civile.

1.2 EVOLUZIONE DELL'APPROCCIO METODOLOGICO PER LA PREVISIONE DELLA CRISI

L'art. 13. co. 2 del Codice della Crisi e dell'Insolvenza d'Impresa assegna al Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili (CNDCEC) il compito di elaborare gli indici necessari al completamento del sistema dell'allerta, introdotto nell'ordinamento con la legge delega n. 155 del 19 ottobre 2017.^[1] La delega contenuta nell'art. 13 CCII affida al CNDCEC il compito di elaborare "gli indici di cui al comma 1 che, valutati unitariamente, consentono una ragionevole presunzione dello stato di crisi."

Dato il richiamo esplicito, è utile ricordare anche il contenuto del primo comma: "Costituiscono indicatori di crisi gli squilibri di carattere reddituale, patrimoniale o finanziario, rapportati alle specifiche caratteristiche dell'impresa e dell'attività imprenditoriale svolta dal debitore, tenuto conto della data di costituzione e di inizio dell'attività, rilevabili attraverso appositi indici che diano evidenza della sostenibilità dei debiti per almeno i sei mesi successivi e delle prospettive di continuità aziendale per l'esercizio in corso o, quando la durata residua dell'esercizio al momento della valutazione è inferiore a sei mesi, per i sei mesi successivi. A questi fini, sono indici significativi quelli che misurano la

sostenibilità degli oneri dell'indebitamento con i flussi di cassa che l'impresa è in grado di generare e l'adeguatezza dei mezzi propri rispetto a quelli di terzi (...)"

Stante l'indicazione normativa, il Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili (CNDCEC) ha svolto preliminarmente una analisi sulla letteratura in materia di modelli di previsione della crisi, riportata di seguito. Secondo questa analisi, lo studio dei modelli di previsione della crisi è divisibile in tre periodi: dal 1930 al 1968, dal 1968 al 1980 e dal 1980 ad oggi. Ciascuno di questi periodi è stato caratterizzato da un approccio metodologico differente e da livelli sempre crescenti di complessità. Nel primo periodo, lo studio della probabilità della insolvenza è stato condotto in funzione di indicatori contabili considerati singolarmente, utilizzando il cosiddetto approccio 'univariato'. Il secondo periodo, iniziato nel 1968 con il primo paper di Altman,^[3] ha visto la diffusione del primo modello multivariato, l'analisi discriminante lineare. Il terzo periodo è stato caratterizzato da modelli probabilistici, in particolare dalla regressione logistica, il cui obiettivo è fornire una misura della probabilità del fallimento di un'impresa. Gli anni dal 1990 ad oggi hanno assistito alla diffusione di modelli di previsione basati sul *machine learning*. Questo breve riassunto permette di contestualizzare l'obiettivo di questa tesi e le motivazioni che la caratterizzano. L'utilizzo di tecniche recenti basate sul cosiddetto *machine*

learning per l'individuazione di schemi ricorrenti in dinamiche sociali come la crisi d'impresa è particolarmente complesso dal punto di vista dell'interpretazione dei risultati ottenuti. Inoltre, il funzionamento di queste tecniche si basa su una quantità e qualità di dati non sempre disponibili nei casi reali. In aggiunta, l'utilizzo del *machine learning* potrebbe portare con sé tempi di addestramento dei modelli e costi computazionali più alti rispetto a quelli di approcci tradizionali. Di conseguenza, in caso di applicabilità di questi modelli, è auspicabile l'ottenimento di una capacità previsiva più elevata.

Capitolo 2

IL MODELLO CNDCEC

Nelle seguenti pagine, verrà inizialmente descritto il modello elaborato dal CNDCEC, il quale si pone come obiettivo la costruzione un sistema di indicatori che permetta di valutare la condizione dello stato giuridico dell'impresa e prevedere un'eventuale situazione di crisi nel prossimo futuro. In seguito verrà mostrata l'applicazione del modello ai dati di imprese estratti dal database di Analisi Informatizzata delle Aziende Italiane (AIDA), gestito dalla società Bureau Van Dijk.

2.1 DESCRIZIONE DEL MODELLO REALIZZATO DAL CNDCEC

Il CNDCEC, pertanto, con il documento pubblicato in data 20 ottobre 2019,^[4] ha proposto una serie di indici. Nello specifico, si tratta di un sistema di indici di tipo gerarchico per il quale gli stessi vanno applicati secondo un dato ordine e uno in subordine dell'altro. Sono stati generati due indici applicabili a tutte le imprese e cinque indici di settore. Il primo indice di crisi, applicabile a tutte le imprese,

viene utilizzato nel caso in cui l'impresa riporti un patrimonio netto negativo, ovvero inferiore ai limiti di legge per le società di capitali. Qualora il patrimonio netto sia positivo, quindi superiore ai limiti di legge, occorre verificare la sostenibilità della situazione debitoria dell'impresa nei successivi sei mesi di esercizio. A tal fine si deve fare riferimento al *debt service coverage ratio* (DSCR), ovvero al rapporto tra tutte le entrate di liquidità previste nei successivi sei mesi e le uscite previste per rimborsi di debiti finanziari nello stesso arco temporale. Nell'eventualità che tale rapporto sia inferiore a uno, è ragionevole presumere lo stato di crisi dell'impresa, purché gli organi di amministrazione e di controllo ritengano affidabili i dati disponibili. Nell'eventualità che il DSCR non sia disponibile o non sia ritenuto affidabile per l'inadeguatezza dei dati a disposizione, trovano applicazione cinque indici con valori soglia differenziati per settori economici, il cui congiunto superamento fa ragionevolmente presumere lo stato di crisi.

I detti cinque indici sono:

1. indice di sostenibilità degli oneri finanziari in termini di rapporto tra gli oneri finanziari ed il fatturato;
2. indice di adeguatezza patrimoniale, in termini di rapporto tra patrimonio netto e debiti totali;

3. indice di ritorno liquido dell'attivo, in termini di rapporto da cash flow e attivo;
4. indice di liquidità, in termini di rapporto tra attività a breve termine e passivo a breve termine;
5. indice di indebitamento previdenziale e tributario, in termini di rapporto tra l'indebitamento previdenziale e tributario e l'attivo.

Per chiarezza espositiva, nel resto dell'elaborato, ci si riferirà a questo gruppo di indici come indici secondari del modello CNDCEC.

In linea con quanto previsto al già menzionato art. 13, il CNDCEC ha altresì previsto specifici indici di crisi per le imprese che si trovano in situazioni che non permetterebbero l'applicazione degli indicatori sopra elencati. A titolo esemplificativo, è stato precisato che per le imprese di nuova costituzione (inferiore ai due anni) assume rilevanza il solo patrimonio netto negativo.

I cinque indici sopracitati ex art. 13. co. 2 sono calcolati a partire da valori presenti all'interno del bilancio e d'esercizio e sono calcolati nel seguente modo:

- Indice di sostenibilità degli oneri finanziari. È costituito dal rapporto tra oneri finanziari e fatturato ed include:

- al numeratore, gli interessi e altri oneri finanziari di cui alla voce C.17 art. 2425 c.c.;
 - al denominatore, i ricavi netti, ovvero la voce A.1) Ricavi delle vendite e prestazioni dell'art. 2425 c.c.
- **Indice di adeguatezza patrimoniale.** È costituito dal rapporto tra il patrimonio netto ed i debiti totali ed include:
 - al numeratore, il patrimonio netto costituito dalla voce A stato patrimoniale passivo dell'art. 2424 c.c., detratti i crediti verso soci per versamenti ancora dovuti (voce A stato patrimoniale attivo) e i dividendi deliberati;
 - al denominatore, i debiti totali costituiti da tutti i debiti (voce D passivo dell'art. 2424 c.c.), indipendentemente dalla loro natura e dai ratei e risconti passivi (voce E passivo dell'art. 2424 c.c.).
 - **Indice di ritorno liquido dell'attivo.** È costituito dal rapporto tra il cash flow e il totale attivo ed include:
 - al numeratore, il cash flow ottenuto come somma del risultato dell'esercizio e dei costi non monetari (ad.es, ammortamenti, svalutazioni crediti, accantonamenti per rischi), dal quale dedurre i

ricavi non monetari (ad.es, rivalutazioni partecipazioni, imposte anticipate);

- al denominatore il totale dell'attivo dello stato patrimoniale art. 2424 c.c.

• **Indice di liquidità.** È costituito dal rapporto tra il totale delle attività ed il totale delle passività a breve termine ed include:

- al numeratore, l'attivo a breve termine quale risultante dalla somma delle voci dell'attivo circolante (voce C attivo dell'art. 2424 c.c.) esigibili entro l'esercizio successivo e i ratei e risconti attivi (voce D attivo dell'art. 2424 c.c.);
- al denominatore, il passivo a breve termine costituito da tutti i debiti (voce D passivo) esigibili entro l'esercizio successivo e dai ratei e risconti passivi (voce E).

• **Indice di indebitamento previdenziale o tributario.** È costituito dal rapporto tra il totale dell'indebitamento previdenziale e tributario ed il totale dell'attivo.

Esso include:

- al numeratore, l'Indebitamento tributario rappresentato dai debiti tributari (voce D.12 passivo dell'art. 2424 c.c.) esigibili entro e oltre

l'esercizio successivo, l'Indebitamento previdenziale costituito dai debiti verso istituti di previdenza e assistenza sociale (voce D.13 passivo dell'art. 2424 c.c.) esigibili entro e oltre l'esercizio successivo;

- al denominatore, l'attivo netto corrispondente al totale dell'attivo dello stato patrimoniale art. 2424 c.c.

Il CNDCEC, come anticipato, ha altresì elaborato un sistema di soglie di questi indicatori differenziato per 10 settori: questi sono un raggruppamento di diverse tipologie di Codici ATECO 2007 come descritto nella Tabella 2.1.

Tabella 2.1: Aggregazioni dei settori secondo le definizioni ATECO 2007

<i>SEZIONE ATECO 2007</i>	<i>GRUPPO O CATEGORIA ATECO 2007</i>	<i>DESCRIZIONE</i>	<i>DEFINIZIONE SETTORE AGGREGATO</i>
<i>A</i>		AGRICOLTURA, SILVICOLTURA E PESCA	(A) AGRICOLTURA SILVICOLTURA E PESCA
<i>B</i>		ESTRAZIONE DI MINERALI DA CAVE E MINIERE	(B) ESTRAZIONE + (C)MANIFATTURA+
<i>C</i>		ATTIVITÀ MANIFATTURIERE	
<i>D</i>	35,11	Produzione di energia elettrica	(D) PROD.ENERGIA/GAS
<i>D</i>	35,21	Produzione di gas	
<i>E</i>		FORNITURA DI ACQUA; RETI FOGNARIE, ATTIVITÀ DI GESTIONE DEI RIFIUTI E RISANAMENTO	(E) FORN.ACQUA RETI FOGNARIE RIFIUTI +
<i>D</i>	35,12	Trasmissione di energia elettrica	
<i>D</i>	35,13	Distribuzione di energia elettrica	(D) TRASM. ENERGIA/GAS
<i>D</i>	35,22	Distribuzione di combustibili gassosi mediante condotte	
<i>D</i>	35,3	Fornitura di vapore e aria condizionata	
<i>D</i>	35,2	PRODUZIONE DI GAS; DISTRIBUZIONE DI COMBUSTIBILI GASSOSI MEDIANTE CONDOTTE	

<i>D</i>	49,5	Trasporto mediante condotte	
<i>F</i>	41	COSTRUZIONE DI EDIFICI	(F41) COSTRUZIONE DI EDIFICI
<i>F</i>	42	INGEGNERIA CIVILE	(F42) INGEGNERIA CIVILE + (F43) COSTR. SPECIALIZZATE
<i>F</i>	43	LAVORI DI COSTRUZIONE SPECIALIZZATI	
<i>G</i>	45	COMMERCIO ALL'INGROSSO E AL DETTAGLIO E RIPARAZIONE DI AUTOVEICOLI E MOTOCICLI	(G45) COMM. INGROSSO e DETT AUTOVEICOLI + (G46) COMM INGROSSO + (D) DISTRIBUZIONE
<i>G</i>	46	COMMERCIO ALL'INGROSSO (ESCLUSO QUELLO DI AUTOVEICOLI E DI MOTOCICLI)	
<i>G</i>	47	COMMERCIO AL DETTAGLIO (ESCLUSO QUELLO DI AUTOVEICOLI E DI MOTOCICLI)	(G47) COMM. DETTAGLIO + I56) BAR e RISTORANTI
<i>G</i>	56	ATTIVITÀ DEI SERVIZI DI RISTORAZIONE	
<i>H</i>		TRASPORTO E MAGAZZINAGGIO	(H) TRASPORTO E MAGAZZINAGGIO + (I55) HOTEL
<i>I</i>	55	HOTEL	
<i>J</i>		SERVIZI DI INFORMAZIONE E COMUNICAZIONE	(J,M,N) SERVIZI ALLE IMPRESE
<i>M</i>		ATTIVITÀ PROFESSIONALI, SCIENTIFICHE E TECNICHE	
<i>N</i>		NOLEGGIO, AGENZIE DI VIAGGIO, SERVIZI DI SUPPORTO ALLE IMPRESE	
<i>P</i>		ISTRUZIONE	(P,Q,R,S) SERVIZI ALLE PERSONE
<i>Q</i>		SANITA' E ASSISTENZA SOCIALE	
<i>R</i>		ATTIVITÀ ARTISTICHE, SPORTIVE, DI INTRATTENIMENTO E	
		DIVERTIMENTO	
<i>S</i>		ALTRE ATTIVITÀ DI SERVIZI	

Le soglie relative ad ogni settore fanno riferimento alle mediane del sottoinsieme delle imprese insolventi, calcolate sul campione di addestramento del CNDCEC composto da circa 568 mila bilanci, corrispondenti a circa 181 mila imprese.^[4] Nella Tabella 2.2, vengono mostrati i valori delle soglie di ogni indice per ciascun settore.

Tabella 2.2: Soglie settoriali

SETTORE	SOGLIE DI ALLERTA				
	Sostenibilità degli oneri finanziari	Adeguatezza patrimoniale	Indice di liquidità	Ritorno liquido dell'attivo	Indebitamento previdenziale e tributario
(A) AGRICOLTURA SILVICOLTURA E PESCA	2,8	9,4	92,1	0,3	5,6
(B) ESTRAZIONE (C) MANIFATTUR. (D) PROD.ENERGI A/GAS	3,0	7,6	93,7	0,5	4,9
(E) FORN. ACQUA RETI FOGNARIE RIFIUTI (D) TRASM. ENERGIA/GAS	2,6	6,7	84,2	1,9	6,5
(F41) COSTRUZIONE DI EDIFICI	3,8	4,9	108,0	0,4	3,8
(F42) INGEGNERIA CIVILE (F43) COSTR. SPECIALIZZAT.	2,8	5,3	101,1	1,4	5,3
(G45) COMM INGROSSO e DETT AUTOVEICOLI (G46) COMM INGROSSO (D) DISTRIB. ENERGIA/GAS	2,1	6,3	101,4	0,6	2,9
(G47) COMM DETTAGLIO (I56) BAR e RISTORANTI	1,5	4,2	89,8	1,0	7,8

<i>(H)</i> TRASPORTO E MAGAZZIN. <i>(I55) HOTEL</i>	1,5	4,1	86,0	1,4	10,2
<i>(JMN) SERVIZI ALLE IMPRESE</i>	1,8	5,2	95,4	1,7	11,9
<i>(PQRS) SERVIZI ALLE PERSONE</i>	2,7	2,3	69,8	0,5	14,6

Per quanto riguarda la lettura critica del modello, occorre fare una precisazione tecnica sul ruolo assunto dagli indici nella previsione della crisi d'impresa. Secondo esplicito orientamento del CNDCEC, la valutazione degli indici selezionati non si è tradotta in uno scoring multivariato di ponderazione degli indici, ma nella contemporanea evidenza di una combinazione di eventi di superamento di soglie di tenuta, la cui emersione congiunta fosse storicamente associata ad elevata probabilità di condurre ad insolvenza. Inoltre, Il modello di allerta prevede che vengano segnalate le imprese che superano la soglia critica per tutti gli indici di bilancio selezionati. Ciò applica il principio della valutazione unitaria prescritta dalla legge;^[1] per cui superare n-1 soglie su n indici non assume la rilevanza richiesta dalla norma. Le analisi statistiche per l'identificazione degli indicatori di squilibrio di carattere reddituale, patrimoniale o finanziario predittivi dell'insolvenza, così come l'identificazione delle soglie di allerta, sono stati realizzati dal CNDCEC con la collaborazione della Centrale dei

Bilanci di Cerved (CeBi). Da ultimo, appare opportuno evidenziare come l'adozione dei suddetti indici non sia obbligatoria per le imprese. Infatti è lo stesso art. 13 co. 3 del Codice della Crisi di Impresa, a prevedere che qualora un'impresa, per via delle proprie specifiche caratteristiche, non ritenga adeguati tali indici, ne debba spiegare in nota integrativa la ragione. Nella stessa sede deve anche indicare i diversi indici che ritiene possano far ragionevolmente presumere la sussistenza del proprio eventuale stato di crisi. In tal caso, tuttavia, un professionista indipendente deve attestarne l'adeguatezza.

2.2 APPLICAZIONE DEL MODELLO CNDCEC

Tutte le operazioni di campionamento e di *preprocessing* dei dati, come anche il calcolo degli indici, sono state eseguite cercando di rimanere il più fedeli possibile alla prassi metodologica utilizzata dal CNDCEC. I dati necessari per la realizzazione dell'analisi desiderata sono stati scaricati tramite esportazione dalla piattaforma AIDA, con l'applicazione di filtri specifici che verranno spiegati in seguito. Partendo dal presupposto della completa disponibilità di dati di bilancio d'esercizio solo fino all'anno 2020, sono stati presi in considerazione i dati relativi all'anno 2017; in questo modo è stato possibile applicare il modello del CNDCEC

per la previsione dell'eventuale crisi d'impresa fino ai 36 mesi successivi su un campione effettivamente completo. L'orizzonte temporale dei tre anni alla data di chiusura del bilancio è lo stesso utilizzato nell'applicazione del modello da parte del CNDCEC stesso.

Il campione di riferimento è costituito da 138.189 imprese, costruito assicurandosi di soddisfare determinate condizioni:

- non di recente costituzione, ovvero che abbiano depositato almeno tre bilanci;
- con attività economica diversa da quella finanziaria ed immobiliare;
- con dimensione superiore a quella definita dalla Direttiva 2013/34/UE² come microimpresa.

² Direttiva del Parlamento europeo e del Consiglio relativa ai bilanci d'esercizio, ai bilanci consolidati e alle relative relazioni di talune tipologie di imprese, recante modifica della direttiva 2006/43/CE del Parlamento europeo e del Consiglio e abrogazione delle direttive 78/660/CEE e 83/349/CEE del Consiglio.

Sono state quindi escluse le imprese secondo la definizione fornita dalla Direttiva 2013/34/UE, ovvero le imprese che alla data di chiusura del bilancio non superano i limiti numerici di almeno due dei tre criteri seguenti:

- a) totale dello stato patrimoniale: 350,000 EUR;
- b) ricavi netti delle vendite e delle prestazioni: 700,000 EUR;
- c) numero medio dei dipendenti occupati durante l'esercizio: 10

Nel nostro caso, sono state filtrate dal campione le imprese che non superavano la soglia stabilita nei criteri b) e c), per la mancata possibilità di impostare un filtro per il quale i risultati ottenuti soddisfino due di tre condizioni specifiche. Inoltre, la combinazione dei criteri b) e c) ha permesso di ottenere un set informativo più ampio rispetto alle altre.

Il procedimento all'interno della piattaforma AIDA volto all'applicazione delle suddette condizioni è stato eseguito tramite l'impostazione dei seguenti filtri:

- Dati finanziari → Principali variabili →
Ricavi delle vendite (migl. EUR): min = 700
- Numero dipendenti: min = 10

- Tipologia di bilancio e disponibilità → Disponibilità di bilanci → Anni con bilanci disponibili: 2014, 2015, 2016, 2017

I dati necessari per l'analisi sono i seguenti:

- Voci dello stato patrimoniale necessarie per il calcolo degli indici;
- Voci del conto economico necessarie per il calcolo degli indici;
- Codice ATECO 2007 delle imprese considerate ai fini del raggruppamento per settori e la conseguente applicazione delle soglie settoriali;

A causa delle limitazioni imposte dalla piattaforma AIDA in termini di quantità di dati scaricabili per singola interazione con il database, è stato necessario scaricare più *file* CSV, in seguito uniti al fine di creare un'unica tabella contenente tutti i valori di bilancio necessari al calcolo degli indici. In tutte le variabili selezionate sono presenti la valuta in cui sono espresse (EUR) e l'anno di riferimento.

Tabella 2.3: Descrizioni delle variabili dello stato patrimoniale

VARIABILE	DESCRIZIONE
<i>CREDITI VERSO SOCI EUR 2017</i>	Crediti relativi ai conferimenti in denaro deliberati, sia in sede di costituzione della società, sia nella fase di aumento di capitale sociale
<i>TOTALE RIMANENZE EUR 2017</i>	Valore dei beni destinati alla vendita, o utilizzati per creare i prodotti da rivendere, che rimangono in giacenza
<i>Crediti a breve EUR 2017</i>	Totale dei crediti esigibili entro l'anno
<i>TOTALE ATTIVITA' FINANZIARIE EUR 2017</i>	Totale delle risorse provenienti da attività della gestione accessoria
<i>TOT. DISPON. LIQUIDE EUR 2017</i>	Somma delle seguenti voci: depositi bancari e postali, assegni, denaro e valori in cassa
<i>RATEI E RISCONTI ATTIVI EUR 2017</i>	Totale di scritture di assestamento per l'attivo dello stato patrimoniale
<i>TOTALE ATTIVO EUR 2017</i>	Totale delle risorse dell'impresa, quindi dello stato patrimoniale
<i>TOTALE PATRIMONIO NETTO EUR 2017</i>	Misura dei mezzi propri investiti dall'imprenditore e/o dai soci nell'impresa
<i>TOTALE DEBITI EUR 2017</i>	Totale del valore delle passività che rappresentano obbligazioni a pagare un ammontare determinato
<i>DEBITI A BREVE EUR 2017</i>	Valore delle passività da pagare entro l'anno
<i>RATEI E RISCONTI PASSIVI EUR 2017</i>	Totale di scritture di assestamento per il passivo dello stato patrimoniale
<i>Debiti Tributari entro EUR 2017</i>	Debiti di natura tributaria da pagare entro l'esercizio
<i>Debiti Tributari oltre EUR 2017</i>	Debiti di natura tributaria sanabili oltre l'esercizio
<i>Istituti previdenza entro EUR 2017</i>	Debiti di natura previdenziale da pagare entro l'esercizio
<i>Istituti previdenza oltre EUR 2017</i>	Debiti di natura previdenziale sanabili oltre l'esercizio
<i>Utile/perdita di esercizio EUR 2017</i>	Utile o perdita generata dall'attività economica dell'impresa

Tabella 2.4: Descrizioni delle variabili del conto economico

VARIABILE	DESCRIZIONE
<i>TOT. VAL. DELLA PRODUZIONE EUR 2017</i>	Somma di tutti gli elementi che hanno contribuito alla produzione economica di un'impresa
<i>Ricavi vendite e prestazioni EUR 2017</i>	Fatturato relativo alla gestione caratteristica
<i>Variazione lavori EUR 2017</i>	Variazione del valore di progetti in corso
<i>COSTI DELLA PRODUZIONE EUR 2017</i>	Tutti i costi che l'impresa deve sostenere per realizzare una determinata produzione di beni e servizi
<i>Amm. Immob. Immat. EUR 2017</i>	Ammortamento delle immobilizzazioni immateriali
<i>Amm. Immob. Mat. EUR 2017</i>	Ammortamento delle immobilizzazioni materiali
<i>Altre svalut. Immob. EUR 2017</i>	Eventuali svalutazioni delle immobilizzazioni
<i>Svalut. crediti EUR 2017</i>	Eventuali svalutazioni dei crediti in possesso
<i>Totale Oneri finanziari EUR 2017</i>	Costi che derivano dall'attività finanziaria dell'impresa

In seguito, è stata esportata la lista delle imprese che hanno subito un evento giuridico attribuibile alla crisi d'impresa nei tre anni successivi, ai fini di replicare la metodologia utilizzata dal CNDCEC. Il procedimento di applicazione del filtro nella piattaforma AIDA è stato il seguente:

- Situazione giuridica → Eventi giuridici:
 - Tipologia di procedura: Fallimento, Concordato preventivo, Liquidazione coatta amministrativa, Liquidazione giudiziale,

Stato di insolvenza, Amministrazione controllata, Concordato fallimentare, Bancarotta semplice.

- Periodo considerato: dal 01/01/2018 al 31/12/2020

Tabella 2.5: Risultati query di ricerca

	<i>TOTALE IMPRESE</i>	<i>NUMERO IMPRESE IN CRISI</i>
<i>Incluse microimprese</i>	1.251.670	6.390
<i>Escluse microimprese</i>	138.189	231

Tramite un'operazione di *join* tra le due tabelle, utilizzando come vincolo di integrità referenziale il codice fiscale, è stato possibile assegnare una *label* dicotomica relativa allo stato giuridico delle imprese. Nello specifico, è stato assegnato il valore uno alle imprese entrate in crisi nel corso dei tre anni successivi, etichettandole come imprese in crisi; è stato assegnato il valore zero alle restanti occorrenze.

Infine, sono stati calcolati gli indici a partire dalle voci di bilancio presenti nella tabella nel seguente modo:

- Indice di Sostenibilità degli oneri finanziari:

$$\frac{\textit{Totale Oneri finanziari}}{\textit{Ricavi vendite e prestazioni} + \textit{Variazione lavori}}$$

- Indice di liquidità:

$$\frac{\text{Totale Rimanenze} + \text{Crediti a breve} + \text{Attività finanziarie} + \text{Tot. Dispon. Liquide} + \text{Ratei e risconti attivi}}{\text{Debiti a breve} + \text{Ratei e risconti passivi}}$$

- Indice di Ritorno liquido dell'attivo:

$$\frac{\text{Tot. Val. della produzione} - \text{Costi della produzione} - \text{Amm. Immob. Mat.} - \text{Amm. Immob. Immat.} - \text{Altre svalut. Immob.} - \text{Svalut. crediti}}{\text{Totale Attivo}}$$

- Indice di Indebitamento previdenziale e tributario:

$$\frac{\text{Debiti Tributari entro} + \text{Debiti Tributari oltre} + \text{Istituti previdenza entro} + \text{Istituti previdenza oltre}}{\text{Totale Attivo}}$$

- Indice di Adeguatezza patrimoniale:

$$\frac{\text{Totale patrimonio netto} - \text{Crediti verso soci} - \text{Dividendi deliberati su utile}}{\text{Totale Debiti} + \text{Ratei e Risconti passivi}}$$

I dividendi deliberati su utile, non presenti nei dati a disposizione, sono stati calcolati in modo approssimativo con la seguente espressione:

$$\begin{aligned} & \text{Totale patrimonio netto (anno } n) + \text{Utile/perdita di esercizio (anno } n) \\ & \quad - \text{Totale patrimonio netto (anno } n + 1) \end{aligned}$$

Questi vengono considerati pari a zero qualora il calcolo fosse risultato negativo. Come è possibile notare, per ottenere un'approssimazione del valore dei dividendi deliberati ricorrendo unicamente ai dati del bilancio d'esercizio, è stato necessario scaricare i valori del patrimonio netto delle

imprese relativi all'anno successivo. Questi dati sono stati inseriti all'interno della tabella principale tramite un'operazione di *join*, sempre considerando il codice fiscale come vincolo di integrità referenziale. Nel primo capitolo si è fatto riferimento al modello CNDCEC come applicazione di una serie di indicatori in modo gerarchico. Il primo di questi indicatori, si basa sul valore del patrimonio netto stesso: partendo dal campione iniziale, quindi, sono state analizzate in via preventiva tutte le imprese con patrimonio netto negativo o inferiori ai minimi legali per le società di capitali³.

Tabella 2.6: Suddivisione imprese con patrimonio netto negativo

	IMPRESE CON PATRIMONIO NETTO NEGATIVO	
	IMPRESE SANE	IMPRESE IN CRISI
VALORI ASSOLUTI	4938	123
PERCENTUALE SUL TOTALE DELLA CLASSE	3,57%	53,2%

Per le imprese con patrimonio netto positivo, è stato necessario confrontare il valore assunto dagli indicatori calcolati con la relativa soglia

³ Secondo l'art. 2327 del c.c. (disposta dall'art. 20 comma 7 del D.L. 91/2014) il capitale minimo per costituire una S.p.a. è di € 50.000; per costituire una S.r.l. ordinaria invece il capitale minimo dovrà essere di €10.000.

settoriale, in modo da verificare il superamento o meno della soglia per prevedere lo stato giuridico dell'impresa. In particolare, come da prassi metodologica, tutti e cinque gli indici devono superare la relativa soglia per far sì che venga previsto lo stato d'insolvenza dell'impresa. Le soglie degli indici sono espresse in termini percentuali; ai fini del confronto diretto, sono stati moltiplicati per 100 i valori degli indici calcolati.

Data la definizione stessa degli indicatori, occorre considerare la correlazione di essi con il rischio, ovvero:

- Indice di sostenibilità degli oneri finanziari: si accende il segnale in caso di valori maggiori o uguali al valore soglia;
- Indice di adeguatezza patrimoniale: si accende il segnale in caso di valori minori o uguali al valore soglia;
- Indice di liquidità: si accende il segnale in caso di valori minori o uguali al valore soglia;
- Indice di ritorno liquido dell'attivo: si accende il segnale in caso di valori minori o uguali al valore soglia;

- Indice di indebitamento previdenziale e tributario: si accende il segnale in caso di valori maggiori o uguali al valore soglia.

Se il denominatore è nullo, ai fini dei segnali di allerta occorre procedere nel seguente modo:

- $(\text{Oneri finanziari})/\text{Ricavi} \cdot 100$
 - Numeratore > 0, Denominatore = 0: Segnale di Allerta ACCESO
 - Numeratore = 0, Denominatore = 0: Segnale di Allerta SPENTO
- $(\text{Pat. Netto})/(\text{Debiti totali}) \cdot 100$
 - Numeratore > 0, Denominatore = 0: Segnale di Allerta SPENTO
 - Numeratore = 0, Denominatore = 0: Segnale di Allerta ACCESO
- $(\text{Attività a breve termine})/(\text{Passività a breve termine}) \cdot 100$
 - Numeratore > 0, Denominatore = 0: Segnale di Allerta SPENTO
 - Numeratore = 0, Denominatore = 0: Segnale di Allerta ACCESO
- $(\text{Cash flow})/\text{Attivo} \cdot 100$
 - Numeratore > 0, Denominatore = 0: Segnale di Allerta SPENTO

- Numeratore = 0, Denominatore = 0: Segnale di Allerta ACCESO
- $(\text{Indebitamento previdenziale} + \text{tributario}) / \text{Attivo} \cdot 100$
 - Numeratore > 0, Denominatore = 0: Segnale di Allerta ACCESO
 - Numeratore = 0, Denominatore = 0: Segnale di Allerta SPENTO

Considerando che le soglie variano per ogni settore, è stato associato ad ogni impresa il proprio settore corrispondente. Questa informazione, assente nel database AIDA, è stata ricavata a partire dai codici ATECO 2007, come visibile nella Tabella 2.1. Per tale scopo, è stato innanzitutto individuato il codice ATECO 2007 per ogni occorrenza: tramite la piattaforma AIDA, è stata scaricata un'ulteriore tabella in cui era presente il codice ATECO 2007 di ogni impresa, unita tramite un'operazione di *join* alla tabella principale utilizzando il codice fiscale come vincolo di integrità referenziale. In seguito, è stata elaborata e implementata una procedura di assegnazione del settore di appartenenza.

Tabella 2.7: Conteggio settoriale delle imprese

SETTORE	STABILI	CRISI
A	1462	0
B-C-DP	38310	40
E-DT	1634	2
F41	3168	6

F42-F43	5965	2
G45-G46-DD	13203	12
G47-I56	10151	5
H-I55	9130	7
J-M-N	12510	12
P-Q-R-S	5575	3

La procedura di confronto tra i valori degli indici e la relativa soglia ha generato cinque nuove colonne, che corrispondono alla predizione di ciascuno dei cinque indicatori in merito allo stato giuridico dell'impresa, e quindi al superamento della soglia o meno. I valori presenti all'interno delle suddette colonne consistono in una variabile dicotomica che assume valore 1 in caso di superamento della soglia e 0 altrimenti. In questo modo, i valori assunti da tutte e cinque le variabili devono essere pari a 1 per definire l'impresa in crisi secondo il modello del CNDCEC. Per applicare correttamente il modello, si è rivelato necessario eliminare tutte quelle occorrenze per cui non è stato possibile calcolare il valore di uno degli indici a causa della presenza di valori nulli nei dati di bilancio. In particolare, è possibile notare nella Tabella 2.8 come la maggior parte delle imprese rimosse dal campione presentavano un valore nullo relativo all'indice dell'indebitamento previdenziale e tributario.

Tabella 2.8: Valori nulli per indice

<i>INDICE</i>	<i>NUMERO DI VALORI NULLI</i>
<i>Indice di Sostenibilità degli oneri finanziari</i>	0
<i>Indice di Ritorno liquido dell'attivo</i>	446
<i>Indice di liquidità</i>	0
<i>Indice di Indebitamento previdenziale e tributario</i>	28908
<i>Indice di Adeguatezza patrimoniale</i>	0

L'elaborazione della procedura ha fornito anche la possibilità di calcolare il numero di imprese che hanno superato un numero definito di soglie, in cui è possibile notare una correlazione tra il numero di segnali accesi e il tasso di default relativo al cluster di imprese considerato. Come visibile nella Figura 2.1 e in Tabella 2.9, è possibile notare l'andamento crescente del tasso di default all'aumentare degli indici in cui venisse superato il suddetto valore critico.



Figura 2.1: Distribuzione del numero di indici superati da ciascuna impresa

Tabella 2.9: Numero di imprese che superano indicatori con tasso di default

<i>NUMERO DI INDICI SUPERATI</i>	<i>NUMERO DI IMPRESE</i>	<i>TASSO DI DEFAULT</i>
0	23817	0,012%
1	42024	0,038%
2	21749	0,133%
3	10351	0,222%
4	2917	0,446%
5	339	1,475%

2.3 RISULTATI OTTENUTI

L'applicazione del modello ha prodotto i risultati presentati nelle Tabelle 2.10 e 2.11, riassunti nelle misure di performance successivamente spiegate.

Tabella 2.10: Matrice di confusione degli indici secondari del modello CNDCEC

<i>VALORI VERI</i>	<i>VALORI PREDETTI</i>	
	<i>STABILI</i>	<i>CRISI</i>
<i>STABILI</i>	100774	334
<i>CRISI</i>	84	5

Tabella 2.11: Misure di performance degli indici secondari del modello CNDCEC

<i>TP</i>	<i>TN</i>	<i>FP</i>	<i>FN</i>	<i>ACCURACY</i>	<i>PRECISION</i>	<i>RECALL</i>	<i>F_MEASURE</i>
5	100774	334	84	99,69%	1,47%	5,62%	2,34%

L'*accuracy* è il numero di dati previsti correttamente su tutti i dati. Più formalmente, è definito come il numero di veri positivi e veri negativi diviso per il numero di veri positivi, veri negativi, falsi positivi e falsi negativi. Questa misura è tra le prime misure più intuitive per valutare la performance di previsione di un modello, riassumendo quante imprese vengono classificate correttamente rispetto a tutte quelle presenti nel campione di riferimento.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision e *recall*, d'altro canto, sono misure che forniscono informazioni più precise sulla correttezza della classificazione. Nello specifico, la *precision* indica la proporzione delle istanze positive corrette tra tutte quelle previste come positive; la *recall*, anche conosciuta come *sensitivity*, indica la frequenza relativa delle istanze realmente positive che sono state identificate correttamente. Infine il cosiddetto *f-measure* è una media armonica delle due misure di *performance* appena citate, tendendo ad assumere il valore minimo tra le due e fornendo quindi una misura sintetica ed affidabile che considera anche gli errori di classificazione e la loro tipologia.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$f\ measure = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Per avere un primo confronto con la prassi metodologica del CNDCEC, sono state considerate diverse combinazioni degli stessi indici e quindi analizzate le *performance* previsive di tali combinazioni, per verificare se un modello più parsimonioso in termini di indicatori utilizzati avesse garantito una previsione più o meno efficace rispetto al modello completo. Sono state analizzate tutte le possibili combinazioni, prendendo in oggetto dai due ai quattro indicatori. Il numero totale è di 25 combinazioni, calcolato come somma di più coefficienti binomiali, ognuno relativo al numero di indicatori considerati rispetto al numero totale di indicatori. In particolare si è ottenuto, mediante il seguente calcolo:

$$Numero\ di\ combinazioni\ totali = \sum_{k=2}^4 \binom{5}{k} = \sum_{k=2}^4 \frac{5!}{k! \cdot (5-k)!} = 25$$

Ai fini del confronto sono state scelte due combinazioni specifiche: quella con il valore di *f-measure* più alto e quella legata al valore di *recall* più alto

(minor numero di falsi negativi) mantenendo un'accuracy superiore al 90%. Esse sono composte rispettivamente dai seguenti indici:

- Combinazione 1: sostenibilità degli oneri finanziari, ritorno liquido dell'attivo, indice di liquidità, adeguatezza patrimoniale;
- Combinazione 2: ritorno liquido dell'attivo, adeguatezza patrimoniale.

Tabella 2.12: Misure di performance di combinazioni scelte

MISURE DI PERFORMANCE	COMBINAZIONE 1	COMBINAZIONE 2
TP	9	30
TN	100464	91741
FP	664	9367
FN	80	59
ACCURACY	99,28%	90,69%
PRECISION	1,38%	0,32%
RECALL	1,01%	33,71%
F_MEASURE	2,42%	0,63%

2.4 OSSERVAZIONI E COMMENTI SUI RISULTATI

L'applicazione di modelli personalizzati derivati in modo diretto dal modello del CNDCEC ci consente di valutare in modo approfondito il funzionamento degli indici secondari come strumento di previsione della crisi d'impresa e sottolineare le criticità ad esso associate. È pressoché immediato

constatare come, tra le combinazioni messe a confronto, il modello del CNDCEC è caratterizzato dal minor numero di imprese segnalate, come anche dal minor numero di imprese segnalate in modo errato. Inoltre, l'*accuracy* del modello è la migliore tra tutte le combinazioni personalizzate testate. Questi risultati parziali forniscono una prima prova concreta del raggiungimento degli obiettivi prefissati da parte del CNDCEC. Tuttavia, rivolgendo l'attenzione verso misure di *performance* che permettono di valutare l'efficacia di una classificazione in maniera più corretta, possono essere portate avanti diverse considerazioni. La problematica principale risiede nel fatto che gli indici secondari siano riusciti a segnalare solamente il 5,6% delle imprese che hanno subito un evento giuridico legato alla crisi d'impresa nei tre anni successivi a quello di riferimento.

L'efficacia di un modello di classificazione dipende dalla capacità del bordo decisionale (i.e. confine della classe nello spazio dimensionale delle *features*) di separare le due classi; esso è sintetizzato nel cosiddetto bordo di classificazione, che viene esplicitato, a seconda del modello, in un valore soglia oppure, ad esempio, in una funzione decisionale (i. e. la funzione di Altman). Nel caso del modello elaborato dal CNDCEC possiamo notare come questo bordo di classificazione sia descritto da due fattori: dalle soglie

settoriali e dal meccanismo di segnalazione del sistema di allerta. Per quanto riguarda le soglie settoriali, nel paragrafo relativo alla descrizione del modello è stato fatto riferimento alla metodologia applicata per l'identificazione di queste soglie. Dal punto di vista tecnico, l'individuazione di un *threshold* in grado di separare due classi quanto più efficacemente possibile parte dallo studio della distribuzione dei valori assunti da una specifica variabile per ognuna delle due classi. Gli approcci statistici mirati alla ricerca del bordo decisionale possono essere di tipo "frequentista" o di tipo "bayesiano", a seconda delle tecniche utilizzate; la differenza tra questi due approcci è il tipo di assunzioni considerate per lo studio delle distribuzioni e della conseguente ricerca del bordo decisionale, oltre che alla diversa 'traduzione' necessaria ad esplicitare il bordo di classificazione dato il bordo decisionale ottimale trovato. Nonostante ciò, entrambi gli approcci hanno come obiettivo l'identificazione del miglior bordo decisionale che separa le classi commettendo il minor numero di errori.

Nel nostro caso, le distribuzioni degli indici non permettono l'utilizzo di questi tipi di approcci in maniera efficace. Come è possibile vedere nelle distribuzioni in Figura 2.2 e in Figura 2.3, non ci sono differenze visivamente

apprezzabili tra i valori assunti dagli indici nelle rispettive classi⁴, che ricordiamo nel nostro caso essere le imprese stabili e le imprese che sono andate in crisi nei 36 mesi successivi all'anno di riferimento dell'indagine condotta. La Tabella 2.13 contiene le statistiche descrittive degli indicatori suddivisi per classe. Per quanto si possono notare alcuni discostamenti tra le medie degli indici delle due classi, l'alta deviazione standard non ne permette un confronto efficace.

Questo è stato confermato applicando modelli di *machine learning* utilizzando gli indici come variabili, con risultati ancor meno performanti del modello del CNDCEC. Il CNDCEC stesso spiega come, inizialmente, sia stato utilizzato un approccio esperto basato sulla costruzione di un albero decisionale⁵. Tuttavia, tale metodologia sembra essersi rivelata sub-ottimale rispetto all'obiettivo prefissato, non consentendo di filtrare in modo adeguato un cluster di insolventi, per giunta coinvolgendo un numero elevato di falsi positivi. Conseguentemente è stato adottato un approccio basato sulla

⁴ Sono state prese come esempio solamente le distribuzioni di 2 dei 5 indici (indice di liquidità e indice di indebitamento previdenziale e tributario) in quanto le distribuzioni degli altri indici si sono presentate ancor meno rilevanti come esempio per il caso di studio.

⁵ Modello di *machine learning* che si pone come obiettivo separare le istanze in classi diverse utilizzando un approccio statistico bayesiano nella costruzione dei bordi di classificazione.

compresenza di più eventi, le cui soglie sono state individuate basandosi sull'osservazione della mediana dell'indice calcolata sui soggetti che hanno presentato nei 36 mesi successivi all'anno di riferimento un evento di default. La mediana, quindi, è stata reputata il *threshold* ottimale ai fini dell'indagine compiuta.

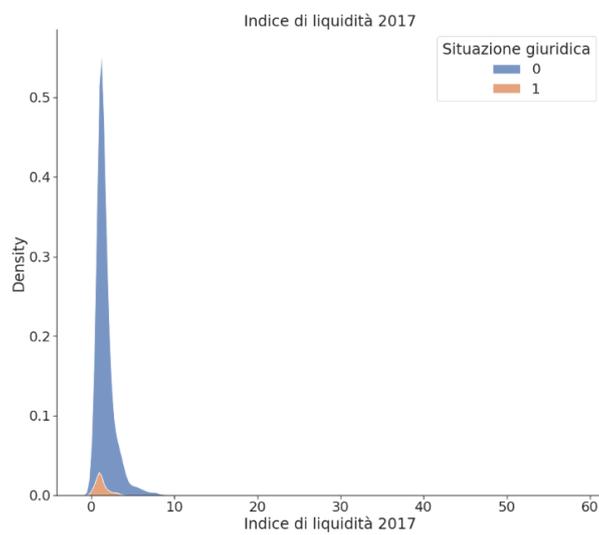


Figura 2.2: Distribuzione dell'indice di liquidità 2017, per classe

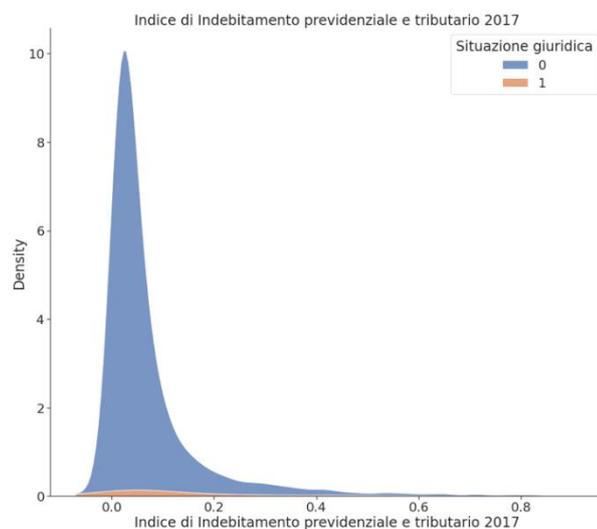


Figura 2.3: Distribuzione dell'indice di indebitamento previdenziale e tributario 2017, per classe

Tabella 2.13: Statistiche descrittive degli indicatori per classe

		SITUAZIONE GIURIDICA	
		0	1
Statistiche Descrittive			
<i>Indice di Sostenibilità degli oneri finanziari</i>	media	1.609	3.456
	std	155.453	9.778
	min	-38.964	0
	25%	0.102	0.321
	50%	0.399	0.903
	75%	1.077	1.993
	max	49110.977	65.400
<i>Indice di Ritorno liquido dell'attivo</i>	media	-0.305	-6.789
	std	22.390	15.344
	min	-5833.699	-95.415
	25%	-4.795	-11.543
	50%	-0.336	-4.154

	75%	3.987	-0.818
	max	427.952	37.656
Indice di liquidità	media	185.594	126.932
	std	1562.549	74.233
	min	0	8.517
	25%	103.148	83.424
	50%	130.999	108.171
	75%	186.141	153.548
	max	425352.756	369.648
Indice di indebitamento previdenziale e tributario	media	7.039	16.607
	std	10.273	19.115
	min	0	0.233
	25%	1.701	2.549
	50%	3.439	6.946
	75%	7.435	25.437
	max	99.497	84.281
Indice di Adeguatezza patrimoniale	media	85.572	12.454
	std	990.781	75.247
	min	-887.170	-625.662
	25%	11.360	3.078
	50%	30.564	10.402
	75%	78.3734	23.459
	max	264656.693	226.555

Se questo significa che non è possibile elaborare strategie ben definite per la definizione di soglie migliori al fine di migliorare le *performance* del modello, è possibile fare una considerazione sull'approccio del CNDCEC definito dai propri obiettivi. Con il confronto tra le diverse combinazioni, si può constatare che per riuscire a individuare un numero più elevato di imprese in crisi e di ridurre il numero di falsi negativi, si ricorre all'inevitabile *trade-off* di aumentare il numero di falsi positivi. La specificazione degli

obiettivi da raggiungere, come anche il peso di questo *trade-off*, dipendono dal costo sociale che si sostiene nel commettere un determinato errore di classificazione. Da un lato, è presente il disagio economico e sociale come conseguenza di un fallimento di un'impresa che potrebbe essere stata salvata in caso di una segnalazione tempestiva. Nel caso opposto, si avrebbe una segnalazione di crisi d'impresa agli organismi di composizione della crisi d'impresa (OCRI), i quali hanno il compito di ricevere le segnalazioni di allerta e gestire il procedimento di allerta nonché di gestire il procedimento di composizione assistita della crisi. Il collegio dispone l'archiviazione delle segnalazioni ricevute se ritiene che non sussista la crisi o che si tratti di imprenditore al quale non si applicano gli strumenti di allerta; evento verosimile nel caso di segnalazioni dei falsi positivi. La stessa segnalazione di crisi dell'impresa potrebbe creare conflitti d'interesse persino tra varie istituzioni presenti all'interno della stessa. Gli organi di controllo dell'impresa⁶ "hanno l'obbligo di verificare che l'organo amministrativo valuti costantemente, assumendo le conseguenti idonee iniziative, se l'assetto organizzativo dell'impresa è adeguato, se sussiste l'equilibrio economico

⁶ il sindaco, il revisore contabile o la società di revisione.

finanziario e qual è il prevedibile andamento della gestione, nonché di segnalare immediatamente allo stesso organo amministrativo l'esistenza di fondati indizi della crisi".^[1] Una possibile criticità potrebbe risiedere proprio nel diverso approccio nella valutazione della gestione aziendale tra gli organi di controllo e l'organo amministrativo dell'impresa.

Queste considerazioni permettono di trarre delle prime conclusioni in merito agli indici secondari scelti nel modello del CNDCEC. Questi sono risultati inefficaci nella previsione della crisi per le imprese fallite entro i tre anni dopo la data di riferimento, per il campione analizzato. Questo perché, dato l'approccio particolarmente conservativo del modello, l'attivazione del segnale di allerta avviene solamente con l'attivazione di tutti e cinque gli indici, limitando il numero totale di segnalazioni come anche, di conseguenza, il numero di segnalazioni errate (i cosiddetti falsi positivi).

Con specifiche combinazioni degli indici, viene ridotto questo approccio conservativo, privilegiando il riconoscimento di potenziali situazioni d'insolvenza con il rischio di commettere più segnalazioni errate. L'utilità degli indici scelti dal CNDCEC, quindi, non risiede tanto nella capacità previsiva della crisi d'impresa quanto nel fornire *insight* di valore riguardo

aree gestionali dell'attività d'impresa, che considerati in modo collettivo possono aiutare nell'identificazione di possibili situazioni di crisi future.

Per una valutazione complessiva del modello elaborato dal CNDCEC, è necessario considerare anche le fasi antecedenti all'applicazione degli indici analizzati fino a questo momento. Il primo indice da applicare secondo il modello riguarda il patrimonio netto, per il quale vengono automaticamente definite in crisi le imprese con patrimonio netto negativo. Come visibile nella Tabella 2.6, più della metà delle imprese entrate in crisi nei 36 mesi successivi sarebbero state segnalate in base all'applicazione di questo indice.

Riassumendo, circa il 97% delle imprese segnalate in crisi dal modello CNDCEC provengono dall'applicazione di questo indice. In questo caso, le segnalazioni dei falsi positivi possono essere giustificate dal fatto che, comunque, si tratterebbero di imprese il cui patrimonio netto sarebbe inferiore al limite previsto dall'art. 2327 del c.c. per le società di capitali.

Il patrimonio netto, come anticipato nella Tabella 2.3, è l'insieme di mezzi propri determinato dalla somma del capitale conferito dal proprietario (o dai soci) in sede di costituzione dell'impresa o durante la vita della stessa con apporti successivi e dall'autofinanziamento. Detto ciò, un valore negativo potrebbe rappresentare una minaccia verosimile alla continuità aziendale

citata nel primo capitolo. L'efficacia dell'indice del patrimonio netto nell'identificare le imprese in crisi verrà utilizzata come spunto per il proseguimento dell'analisi, fornendo una base teorica ed interpretativa per la costruzione di un modello personalizzato. Per riassumere l'analisi condotta finora, nella Tabella 2.14 verranno mostrate le *performance* complessive del modello CNDCEC nel suo insieme.

Tabella 2.14: Confusion matrix e performance complessive del modello CNDCEC

<i>VALORI VERI</i>	<i>VALORI PREDETTI</i>	
	<i>STABILI</i>	<i>CRISI</i>
<i>STABILI</i>	100774	5272
<i>CRISI</i>	84	128

<i>MISURE DI PERFORMANCE</i>	<i>MODELLO CNDCEC COMPLETO</i>
<i>ACCURACY</i>	94,96%
<i>PRECISION</i>	2,37%
<i>RECALL</i>	60,38%
<i>F_MEASURE</i>	4,56%

Capitolo 3

MACHINE LEARNING AI FINI DELLA PREVISIONE DELLA CRISI

In questo capitolo, partendo dalle considerazioni fatte nelle pagine precedenti, verranno presentati i problemi intrinseci riguardanti la previsione della crisi d'impresa. In seguito, verrà descritto l'utilizzo del *machine learning*, dimostrando il percorso logico per la motivazione della scelta della tecnica che sarà implementata. Infine ne verrà descritto il funzionamento e ne saranno riportati e commentati i risultati ottenuti.

3.1 CRITICITÀ NELLA PREVISIONE DELLA CRISI D'IMPRESA

Per quanto la previsione della crisi d'impresa possa essere considerata di notevole importanza, porsi come obiettivo la costruzione di un modello volto a prevedere il futuro stato d'insolvenza di un'azienda porta con sé diverse criticità.

In primis, è da considerare il fatto che la crisi, in quanto tale, è un evento non comune. Di conseguenza, anche nel campione di dati utilizzato per l'analisi, il numero di imprese che si sono trovate in uno stato di crisi è drasticamente inferiore a quello delle imprese sane. Quando si è in presenza di classi così sbilanciate (imprese stabili e imprese in crisi) è necessario tenere conto dei seguenti aspetti:

- La classe minoritaria non possiede un numero sufficiente di istanze tale da permettere l'individuazione di un pattern caratteristico della classe stessa;
- I modelli di classificazione basati sul *machine learning* verrebbero 'addestrati' su dati sbilanciati, risultando in un più facile riconoscimento delle istanze della classe maggioritaria.

Il primo aspetto riguarda l'attendibilità dei risultati che si ottengono da un qualsiasi tipo di analisi che viene condotta su una quantità inadeguata di dati. In questo caso, trattandosi di dati relativi a voci di bilancio di diverse realtà aziendali, una caratteristica di primo rilievo è la variabilità dei valori assunti dalle *features*, che rende ancora più importante la considerazione di questa problematica.

Il secondo aspetto è ancor più rilevante poiché l'addestramento di questo tipo di modelli si basa sulla minimizzazione dell'errore di classificazione: a fronte di un numero di elementi più elevato di una classe, il modello tende a privilegiare la classificazione corretta degli elementi della classe maggioritaria rispetto all'altra.

Per poter ovviare alle problematiche sopra descritte esistono diverse soluzioni. Ad esempio, si potrebbe ricorrere all'utilizzo dell'*undersampling*, un approccio secondo il quale viene ridotto il numero di istanze della classe maggioritaria al fine di ottenere un bilanciamento tra le classi: così facendo, però, si andrebbe incontro ad una perdita di informazione rilevante. Il campione della classe maggioritaria scelta, una volta ridotto, potrebbe non rappresentare accuratamente la realtà e il risultato dell'analisi potrebbe essere impreciso. Alternativamente, ci si potrebbe avvalere di una tecnica di *oversampling* denominata SMOTE (*Sinthetic Minority Oversampling Technique*) che, considerando i dati a disposizione come dei punti in uno spazio n-dimensionale (con n numero delle *features*), si pone come obiettivo quello di bilanciare le classi creando delle istanze fittizie della classe minoritaria (vicino ai punti esistenti).^[5] Tenuto conto però del forte sbilanciamento dei dati nel caso di studio approfondito nelle pagine precedenti, seguire questo metodo

significherebbe lavorare con un numero eccessivo di istanze sintetiche. Tuttavia, pur considerando l'alternativa di un approccio misto tra l'*undersampling* della classe maggioritaria e l'*oversampling* della minoritaria, al fine di ridurre il numero di istanze sintetiche totali, quest'ultimo porta con sé criticità dal punto di vista interpretativo. La creazione di istanze fittizie si tradurrebbe nella considerazione di dati realistici ma non reali, il che non è stato ritenuto opportuno in quanto porterebbe alla creazione di imprese con annessi dati di bilancio.

Un altro aspetto che rende critica la previsione della crisi d'impresa è l'elevata correlazione tra le variabili di bilancio. I valori che assumono le voci di bilancio di un'impresa derivano dalla gestione dell'attività economica, che consiste in una serie di processi aziendali integrati. A causa di questa integrazione, è naturale conseguenza notare un certo livello di correlazione tra le diverse variabili presenti nel bilancio di una società.⁷

Dal punto di vista tecnico, però, la presenza di *features* fortemente correlate potrebbe creare problemi nello sviluppo di modelli previsivi a causa di assunzioni matematiche e statistiche che non verrebbero pienamente soddisfatte. Ad

⁷ I valori di tale correlazione per le principali voci di bilancio sono reperibili nella Tabella A.1 dell'appendice.

esempio, per i modelli lineari (e.g. regressione lineare o regressione logistica), la correlazione tra le variabili indipendenti (i.e. multicollinearità) può condurre a soluzioni numericamente instabili e, dunque, a un'inattendibilità dei risultati ottenuti.

Secondo una visione interpretativa, la presenza di variabili tra loro correlate non arricchirebbe il contenuto informativo in quanto queste sono, per definizione, simili. Facendo riferimento al modello del CNDCEC, l'utilizzo di indici, per quanto non utilizzati all'interno di un modello di *machine learning*, permette di far fronte alla problematica della multicollinearità in quanto questi racchiudono in singole variabili l'efficacia di determinati processi e attività aziendali, sacrificando tuttavia parte dell'informazione.⁸

3.2 ANALISI DISCRIMINANTE LINEARE

Alla luce dell'inadeguatezza della seconda parte del modello CNDCEC nella previsione dell'eventuale crisi d'impresa, si è optato per l'utilizzo di un modello di

⁸ I valori di tale correlazione per gli indici secondari del CNDCEC sono reperibili nella Tabella A.1 dell'appendice.

machine learning auspicando una migliore performance. La scelta è ricaduta nell'utilizzo dell'analisi discriminante lineare (o LDA) come algoritmo ai fini della creazione di un modello di classificazione. Le motivazioni della scelta sono molteplici:

- La LDA consente la cosiddetta *feature extraction*, creando una combinazione lineare delle variabili considerate al fine di ottenerne una nuova che discrimina al meglio le istanze tra le classi;
- è un algoritmo robusto in quanto consente di ottenere performance elevate anche in caso di violazione di ipotesi matematiche e statistiche, come quella della correlazione tra le variabili;
- in letteratura, sono esistenti casi di applicazione di LDA che hanno riscontrato esiti positivi, come lo Z-Score di Altman^[3] e le sue successive varianti.

La LDA funziona calcolando statistiche di riepilogo per ciascuna variabile per entrambe le classi, come la media e la deviazione standard. Nello specifico, viene utilizzata la scomposizione di matrici per calcolare le statistiche in modo efficiente; queste rappresentano il modello appreso dai dati di addestramento. Le previsioni vengono effettuate stimando la probabilità che un nuovo esempio appartenga a ciascuna etichetta di classe in base ai valori di ciascuna variabile. La

classe che risulta con la probabilità maggiore viene quindi assegnata all'esempio. In quanto tale, la LDA può essere considerata un'applicazione del teorema di Bayes per la classificazione.^[6]

Prima dello sviluppo del modello di classificazione si è scelto di utilizzare la tecnica di *undersampling* per diminuire la *class imbalance* all'interno del dataset. Il numero di imprese mantenute è stato scelto affinché lo sbilanciamento tra le classi dei dati a disposizione rispecchiasse quello dei dati utilizzati dal CNDCEC per l'addestramento del proprio modello (3,1% di casi di insolvenza).^[4] Sono state quindi considerate un totale di 4.100 imprese stabili e 131 episodi di crisi d'impresa. La riduzione del numero di imprese insolventi rispetto al campione iniziale deriva dalla rimozione delle istanze con valori nulli, le quali avrebbero reso inapplicabile la LDA. I valori delle variabili, inoltre, sono stati standardizzati.

Per verificare l'efficacia di un modello di classificazione è necessario testare quest'ultimo su un campione di dati non utilizzati per l'addestramento (*test set*). In questo modo viene valutata la capacità del modello di classificare correttamente nuove istanze. Il modello è stato addestrato con una *stratified 10-fold cross-validation*. Questa consiste nell'esecuzione ripetuta dell'algoritmo (in questo caso LDA) su campioni di dati, denominati dati di *training*, di volta in volta diversi. In particolare, i dati a disposizione vengono divisi in 10 parti (*fold*),

ciascuna bilanciata come il dataset di partenza. Il modello viene addestrato su un campione di training composto da 9 dei 10 *fold* e testato sui dati rimanenti. La procedura viene ripetuta fino a che ogni *fold* non è stato utilizzato sia come *train set* sia come *test set*. L'utilizzo di questa tecnica di addestramento permette di ottenere un modello che evita l'eccessivo adattamento ai dati di *training* (fenomeno denominato *overfitting*). La *performance* del modello di classificazione addestrato con *cross-validation* si ottiene dalla media delle *performance* dei modelli derivanti da ogni iterazione.

3.3 COMMENTO DEI RISULTATI OTTENUTI E CONCLUSIONI

Nella Tabella 3.1 è possibile notare come le *performance* di classificazione del modello personalizzato siano superiori a quelle del modello CNDCEC.

Tabella 3.1: Confronto tra performance del modello CNDCEC e modello LDA

<i>MISURE DI PERFORMANCE</i>	<i>MODELLO LDA</i>	<i>MODELLO CNDCEC</i>
<i>ACCURACY</i>	96,83%	94,96%
<i>F_MEASURE</i>	24,83%	4,56%

Trattandosi di un problema a due classi, la *Linear Discriminant Analysis* restituisce una sola discriminante lineare, ovvero la variabile ottenuta dalla combinazione lineare delle *features* esistenti; questa combinazione permette di discriminare al meglio le istanze tra le classi. In Tabella 3.2 sono visibili i coefficienti associati a ciascuna delle variabili nel calcolo della discriminante lineare.

Tabella 3.2: Coefficienti della combinazione lineare della LDA

	WEIGHT
Bias	-5,339169263
CREDITI VERSO SOCI	-0,25521425
RATEI E RISCONTI ATTIVI	0,459354711
TOTALE PATRIMONIO NETTO	-8,589944366
TOTALE DEBITI	27,04901051
DEBITI A BREVE	-10,02688249
RATEI E RISCONTI PASSIVI	-0,68669158
Debiti Tributari entro	2,523792713
Debiti Tributari oltre	0,407486211
Istituti previdenza entro	0,571373264
Istituti previdenza oltre	0,813049815
Utile/perdita di esercizio	1,202400656
TOT. VAL. DELLA PRODUZIONE	-73,09076459
COSTI DELLA PRODUZIONE	70,41839513
Amm. Immob. Immat.	-1,361667926
Amm. Immob. Mat.	-0,124052533
Altre svalut. Immob.	0,381328097
Svalut. crediti	0,388892
Totale Oneri finanziari	-14,20041804
TOTALE CREDITI	3,008036953

L'assegnazione dall'etichetta di classe ad una determinata impresa dipende dal valore assunto dalla combinazione lineare considerando i valori di bilancio dell'impresa stessa. Il risultato è da interpretare come la probabilità di assegnare all'impresa lo stato di crisi (classe 1); di conseguenza, il singolo coefficiente rappresenta l'impatto che la corrispondente voce di bilancio assume nel calcolo della probabilità che un'impresa entri in uno stato di crisi nei successivi tre anni.

Focalizzando l'attenzione sui coefficienti della Tabella 3.2, è possibile notare come ci siano coefficienti nettamente maggiori di altri in valore assoluto. Nello specifico, è interessante sottolineare che i costi della produzione e il totale dei debiti abbiano un impatto significativo nel calcolo della probabilità di insolvenza. Al contrario, il valore della produzione e il patrimonio netto influenzano fortemente la riduzione della stima di questa probabilità. Questi risultati sembrano essere in accordo con la natura stessa della voce di bilancio. D'altro canto sono presenti dei coefficienti il cui segno è incoerente con la percezione reale delle variabili ad esse associate, come quello relativo agli oneri finanziari.

Mettendo a confronto le *performance* del modello ottenuto con la LDA con quelle del modello del CNDCEC (complessive), come visibile nella Tabella 3.1, si può notare che la LDA è in grado di fare previsioni più accurate. Nonostante ciò, quest'ultimo modello solleva dei dubbi interpretativi, come nel caso

dell'incoerenza dei coefficienti. Inoltre, la combinazione lineare di diverse voci di bilancio non fornisce un indicatore espressivamente chiaro come gli indici utilizzati dal CNDCEC; tuttavia, applicare la LDA su indici precalcolati è risultato inefficace. Nello specifico, è stata provata l'applicazione della LDA sugli indici utilizzati dal CNDCEC, sul Margine EBITDA e sull'indice di struttura primario.

Al fine di ottenere un modello parsimonioso, da utilizzare come primo confronto con il modello principale in cui vengono considerate tutte le variabili, è stata applicata una tecnica di *step-wise feature selection*: la *recursive feature elimination* (RFE). Dato un algoritmo che assegna pesi alle variabili (nel nostro caso, i coefficienti del modello lineare), l'obiettivo della RFE è quello di selezionare le variabili considerando ricorsivamente insiemi di *features* sempre più piccoli. Partendo dal modello completo, viene eliminata la variabile con il peso minore; tale procedura viene ripetuta in modo ricorsivo sul gruppo di variabili rimanenti fino al raggiungimento del numero di *features* desiderate da selezionare. Ridurre il numero di variabili utilizzate nel modello, consente di semplificarlo, facilitandone la comprensione e la successiva interpretazione. Grazie all'utilizzo di questa tecnica è stato possibile ottenere un modello più parsimonioso con delle *performance* migliori del modello iniziale, come visibile nella Tabella 3.3.

Tabella 3.3: Performance del modello a seguito della RFE

<i>MISURE DI PERFORMANCE</i>	<i>MODELLO DERIVANTE DALLA RFE</i>
<i>ACCURACY</i>	96,90%
<i>F_MEASURE</i>	25,61%

Tabella 3.4: Variabili e pesi del modello derivante dalla LDA con RFE

	<i>WEIGHT</i>
<i>Bias</i>	-5,29004
<i>RATEI E RISCONTI ATTIVI</i>	0,4749
<i>TOTALE PATRIMONIO NETTO</i>	-9,10989
<i>TOTALE DEBITI</i>	20,04525
<i>DEBITI A BREVE</i>	-8,50724
<i>RATEI E RISCONTI PASSIVI</i>	-0,63379
<i>Debiti Tributari entro</i>	2,493295
<i>Debiti Tributari oltre</i>	0,329468
<i>Istituti previdenza entro</i>	0,563796
<i>Istituti previdenza oltre</i>	0,804228
<i>TOT. VAL. DELLA PRODUZIONE</i>	-75,8016
<i>COSTI DELLA PRODUZIONE</i>	73,01082
<i>Amm. Immob. Immat.</i>	-1,01121
<i>Svalut. crediti</i>	0,514481
<i>Totale Oneri finanziari</i>	-8,02592
<i>TOTALE CREDITI</i>	3,785508

Il modello migliore è risultato essere quello con 15 delle 19 variabili iniziali, le quali sono visibili in Tabella 3.4; i coefficienti restano simili a quelli della

Tabella 16. Un'ulteriore riduzione del numero di variabili considerate si è dimostrata essere non altrettanto performante, restituendo delle misure di *performance* peggiori di quelle del modello completo.

CONCLUSIONI E SVILUPPI FUTURI

Dall'analisi svolta nelle precedenti pagine sono emerse una serie di conclusioni.

Innanzitutto, è stata verificata, in un'ottica di classificazione, una relativa incapacità degli indici secondari del modello CNDCEC nel prevedere la crisi d'impresa. Tuttavia questi indici, nella prassi applicativa, assumono un ruolo subordinato rispetto agli altri indicatori all'interno del modello, e sono stati scelti e calcolati seguendo l'obiettivo di minimizzare il numero di segnalazioni errate, il quale è stato raggiunto.

Il modello di *machine learning* sviluppato e presentato nel Capitolo 3 ha permesso di ottenere una *performance* di classificazione superiore al modello CNDCEC, sacrificando tuttavia la capacità interpretativa dei risultati. Questo perché la previsione della crisi d'impresa si sarebbe basata sull'elaborazione della macchina dei valori di bilancio dell'impresa stessa, senza chiarire delle soglie specifiche interpretabili e riferibili ad ambiti di gestione aziendale.

La questione principale, quindi, risiede nella difficoltà di elaborare modellistiche performanti senza rinunciare alla complessità e alla derivante

difficoltà di interpretazione. Questa difficoltà è causata da vari aspetti, sia giuridici che tecnici.

Il primo di questi è attribuibile alla prescrizione di legge. Nel Capitolo 1, citando l'articolo 13 del CCII, è stato fatto riferimento alla necessità di valutare unitariamente il sistema di indici per la segnalazione della crisi. Il fatto che gli indici di cui al comma 2 siano suscettibili di una valutazione unitaria implica non tanto la predisposizione di una misura di sintesi, che da sola indichi la presenza o meno di indizi di crisi (e.g. lo Z-Score di Altman), quanto lo sviluppo di un percorso logico di indici che porti infine alla valutazione unitaria. Inoltre, questo percorso logico deve comprendere indici che siano attribuibili a specifiche aree di gestione aziendale, in modo che ne valutino efficacemente la prestazione essendo singolarmente interpretabili.^[1]

Il secondo aspetto da considerare riguarda il numero di precauzioni tecniche da soddisfare per eseguire un'analisi efficace del tema della crisi d'impresa e della sua previsione. Questi accorgimenti sono stati approfonditi nel Capitolo 3, in cui sono stati dimostrati i limiti e le potenzialità derivanti dall'applicazione di tecniche di *machine learning*. Nello specifico, è stata rilevata l'inadeguatezza di vari indici nel prevedere la crisi d'impresa.

Risulta chiaro che le prescrizioni di legge così definite, hanno l'effetto di limitare il numero di approcci considerabili ai fini della previsione della crisi. Persino il CNDCEC, utilizzando un campione decisamente più ampio e passando in rassegna un elevato numero di indici,^[4] non è stato in grado di ottenere un sistema di indicatori più efficace di quello attuale. Questo perché il vincolo del Legislatore non ha permesso altre strade se non quella di ricercare delle soglie per ogni indicatore utilizzato che permettessero unitariamente di discriminare le imprese stabili dalle imprese in crisi: approccio che si è rivelato inadeguato. La specificazione del bordo di classificazione può essere di vario tipo, ed è fortemente legato alla modello di apprendimento adoperato. Lo Z-Score di Altman, ad esempio, è stato storicamente considerato come un modello di valutazione della situazione aziendale di discreto successo, pur non permettendo una valutazione singolare degli indici ma solamente d'insieme ai fini della classificazione in imprese stabili ed in crisi.

In conclusione, per poter elaborare un modello che sia capace di prevedere in modo efficace una realtà complessa come la crisi d'impresa, è necessario rinunciare alla necessità di dover valutare i singoli componenti del modello secondo un approccio tradizionale. Questo tema potrebbe rappresentare, per il Legislatore, una possibile sfida nel prossimo futuro, in quanto andrebbe a

modificare gli equilibri della valutazione critica di un *topic* delicato come la crisi d'impresa tra il giudizio di un modello e quello degli organi di controllo aziendale.

Sicuramente, l'utilizzo di indici è di fondamentale importanza per garantire un sufficiente livello interpretativo del modello. Le più recenti tecniche, unite alle sempre crescenti capacità di calcolo, rendono possibile la creazione di modelli di classificazione e di previsione della crisi performanti a partire da indici riguardanti gli aspetti aziendali e il sistema economico in cui è collocata l'impresa.^[7]

Un inedito terreno di ricerca potrebbe essere rappresentato dall'applicazione della *anomaly detection*. L'applicazione della LDA sulle voci di bilancio, pur non consentendo l'interpretazione delle aree gestionali dell'impresa, ha permesso di notare un'influenza considerevole delle voci stesse nell'identificazione dell'episodio della crisi. Un'ipotesi potrebbe essere quella di condurre un'analisi sull'andamento di queste voci, verificando se eventuali comportamenti anomali possano essere correlati a casi di difficoltà riguardanti la gestione dell'attività aziendale. La segnalazione tempestiva di tali anomalie potrebbe tradursi nella possibilità di reagire tempestivamente alle cause ad esse associate, riuscendo così a prevedere ma soprattutto a prevenire eventuali situazioni critiche che, aggravandosi, possono sfociare nella crisi d'impresa.

Bibliografia

1. DECRETO LEGISLATIVO 12 gennaio 2019, n. 14. *Codice della crisi d'impresa e dell'insolvenza* in attuazione della LEGGE 19 ottobre 2017, n. 155.

2. Quagli, A. *Il concetto di crisi d'impresa come incontro tra la prospettiva aziendale e quella giuridica*. Crisi d'impresa e fallimento. 2016.

3. *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*. Altman, Edward I. 4, s.l.: Wiley, 1968, The Journal of Finance, Vol. 23, p. 589-609.

4. CNDCEC (Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili). *Crisi d'Impresa - Gli indici di allerta*. 2019.

5. *SMOTE: Synthetic Minority Over-sampling Technique*. N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer. 2002, Journal Of Artificial Intelligence Research, Vol. 16, p. 321-357.

6. Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar. *Introduction to Data Mining (second edition)*. s.l. : Pearson, 2019.

7. *A novel two-stage hybrid default prediction model with k-means*. Kunpeng Yuan, Guotai Chi, Ying Zhou, Hailei Yin. 2022, Research in International and Business Finance, Vol. 59.

APPENDICE

Tabella A.1: Matrice di correlazione delle voci di bilancio⁹

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V1	1.000	0.004	0.007	0.014	0.015	0.005	-0.002	0.010	0.010	0.004	0.019
V2	0.004	1.000	0.293	0.408	0.177	0.607	0.158	0.279	0.253	0.741	0.221
V3	0.007	0.293	1.000	0.600	0.403	0.230	0.598	0.700	0.411	0.538	0.683
V4	0.014	0.408	0.600	1.000	0.912	0.364	0.429	0.536	0.521	0.642	0.850
V5	0.015	0.177	0.403	0.912	1.000	0.189	0.304	0.474	0.473	0.343	0.791
V6	0.005	0.607	0.230	0.364	0.189	1.000	0.171	0.246	0.219	0.549	0.221
V7	-0.002	0.158	0.598	0.429	0.304	0.171	1.000	0.311	0.273	0.282	0.539
V8	0.010	0.279	0.700	0.536	0.474	0.246	0.311	1.000	0.998	0.405	0.734
V9	0.010	0.253	0.411	0.521	0.473	0.219	0.273	0.998	1.000	0.382	0.724
V10	0.004	0.741	0.538	0.642	0.343	0.549	0.282	0.405	0.382	1.000	0.479
V11	0.019	0.221	0.683	0.850	0.791	0.221	0.539	0.734	0.724	0.479	1.000

⁹ Per visualizzare le voci di bilancio associate alle abbreviazioni, riferirsi alla Tabella A.2.

Tabella A.2: Voci di bilancio associate alle abbreviazioni in Tabella A.1

V1	CREDITI VERSO SOCI EUR 2017
V2	RATEI E RISCONTI ATTIVI EUR 2017
V3	TOTALE PATRIMONIO NETTO EUR 2017
V4	TOTALE DEBITI EUR 2017
V5	DEBITI A BREVE EUR 2017
V6	RATEI E RISCONTI PASSIVI EUR 2017
V7	Utile/perdita di esercizio EUR 2017
V8	TOT. VAL. DELLA PRODUZIONE EUR 2017
V9	COSTI DELLA PRODUZIONE EUR 2017
V10	Totale Oneri finanziari EUR 2017
V11	TOTALE CREDITI EUR 2017

Tabella A.3: Matrice di correlazione degli indici secondari del modello

CNDCEC

	Indice di Sostenibilità degli oneri finanziari 2017	Indice di liquidità 2017	Indice di Ritorno liquido dell'attivo 2017	Indice di Indebitamento previdenziale e tributario 2017	Indice di Adeguatezza patrimoniale 2017
Indice di Sostenibilità degli oneri finanziari 2017	1.000	-6.82E-10	-9.52E-11	-7.37E-11	-2.08E-12
Indice di liquidità 2017	-6.82E-10	1.000	0.012	-0.008	0.841
Indice di Ritorno liquido dell'attivo 2017	-9.52E-11	0.012	1.000	-0.634	0.010
Indice di Indebitamento previdenziale e tributario 2017	-7.37E-11	-0.008	-0.634	1.000	-0.015
Indice di Adeguatezza patrimoniale 2017	-2.08E-12	0.841	0.010	-0.015	1.000