



UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA GIORGIO FUÀ

Corso di Laurea Magistrale in Data Science per l'Economia e le Imprese

Analisi dimensionale della redditività d'impresa: il caso del
settore alimentare italiano

Dimensional analysis of business profitability: the case of
the Italian food sector

Relatore:

Prof. Domenico Potena

Tesi di Laurea di:

Chiara Antognozzi

Anno Accademico 2021 - 2022

Indice

1	Introduzione	14
2	Capitolo 1	16
2.1	Fonte del dataset	16
2.2	Business understanding	17
2.2.1	Analisi dell'industria alimentare italiana	17
2.3	Data understanding	20
2.3.1	Struttura del dataset	24
2.3.2	Variabili continue	25
2.3.3	Variabili categoriche	32
2.4	Preprocessing	35
2.4.1	Data Cleaning	35
2.4.2	Data Integration	36
2.4.3	Data Normalization	39
3	Capitolo 2	42

3.1	Cluster	42
3.2	DBSCAN	43
3.2.1	Definizione	43
3.2.2	Funzionamento	44
3.2.3	Metriche	45
3.2.4	Analisi	46
3.3	K-MEANS	49
3.3.1	Funzionamento	49
3.3.2	Metriche	50
3.3.3	Analisi	50
4	Capitolo 3	56
4.1	Composizione del dataset	56
4.2	Random Forest per Features Selection	58
4.2.1	Funzionamento	58
4.2.2	Metriche	59
4.2.3	Analisi	60
4.3	Loop Attributes Subset	63
4.4	Decision Tree	65
4.4.1	Funzionamento	65
4.4.2	Analisi	66
4.5	Random Forest	79

4.5.1	Analisi	79
4.6	Support Vector Regression	92
4.6.1	Funzionamento	92
4.6.2	Analisi	94
5	Capitolo 4	109
5.1	Risultati	109
5.1.1	Piccole Imprese	110
5.1.2	Medie Imprese	111
5.1.3	Grandi Imprese	112
5.1.4	Discussione	113
6	Conclusioni	117

Elenco delle tabelle

2.1	Caratteristiche delle microimprese	18
2.2	Caratteristiche delle microimprese.2	19
2.3	Struttura del dataset	24
2.4	Distribuzioni delle variabili in migliaia di euro	25
2.5	Distribuzione delle nuove variabili macroeconomiche	38
2.6	Range delle variabili standardizzate	41
3.1	Setting dei parametri del DBSCAN	47
3.2	Setting dei parametri del K-MEANS	50
3.3	Caratteristiche dei cluster	52
3.4	Caratteristiche dei cluster	54
3.5	Caratteristiche dei cluster	55
4.1	Training set	56
4.2	Test set	57
4.3	Composizione del Training e Test set	57

4.4	PICCOLE IMPRESE	60
4.5	MEDIE IMPRESE	61
4.6	GRANDI IMPRESE	61
4.7	Nuovi cluster	64
4.8	Setting dei parametri del Decision Tree	66
4.9	Tuning dei parametri del Decision Tree: piccole imprese Spa Nord	66
4.10	Tuning dei parametri del Decision Tree: piccole imprese Spa Centro	67
4.11	Tuning dei parametri del Decision Tree: piccole imprese Spa Sud/Isole	67
4.12	Tuning dei parametri del Decision Tree: piccole imprese Srl Nord	68
4.13	Tuning dei parametri del Decision Tree: piccole imprese Srl Centro	68
4.14	Tuning dei parametri del Decision Tree: piccole imprese Srl Sud/Centro	69
4.15	Tuning dei parametri del Decision Tree: piccole imprese Società Cooperative Nord	69
4.16	Tuning dei parametri del Decision Tree: piccole imprese Società Cooperative Centro	70
4.17	Tuning dei parametri del Decision Tree: piccole imprese Società Cooperative Sud/Isole	70

4.18	Tuning dei parametri del Decision Tree: medie imprese Spa Nord	71
4.19	Tuning dei parametri del Decision Tree: medie imprese Spa Centro	71
4.20	Tuning dei parametri del Decision Tree: medie imprese Spa Sud/Isole	72
4.21	Tuning dei parametri del Decision Tree: medie imprese Srl Nord	72
4.22	Tuning dei parametri del Decision Tree: medie imprese Srl Centro	73
4.23	Tuning dei parametri del Decision Tree: medie imprese Srl Sud/Isole	73
4.24	Tuning dei parametri del Decision Tree: medie imprese Società Cooperative Nord	74
4.25	Tuning dei parametri del Decision Tree: medie imprese Società Cooperative Centro	74
4.26	Tuning dei parametri del Decision Tree: medie imprese Società Cooperative Sud/Isole	75
4.27	Tuning dei parametri del Decision Tree: grandi imprese Spa Nord	75
4.28	Tuning dei parametri del Decision Tree: grandi imprese Spa Centro	76
4.29	Tuning dei parametri del Decision Tree: grandi imprese Spa Sud/Isole	76
4.30	Tuning dei parametri del Decision Tree: grandi imprese Srl Nord	77
4.31	Tuning dei parametri del Decision Tree: grandi imprese Srl Centro	77

4.32 Tuning dei parametri del Decision Tree: grandi imprese Srl	
Sud/Isole	78
4.33 Tuning dei parametri del Decision Tree: grandi imprese Società	
Cooperative Nord	78
4.34 Setting dei parametri del Random Forest	79
4.35 Tuning dei parametri del Random Forest: piccole imprese Spa	
Nord	80
4.36 Tuning dei parametri del Random Forest: piccole imprese Spa	
Centro	80
4.37 Tuning dei parametri del Random Forest: piccole imprese Spa	
Sud/Isole	81
4.38 Tuning dei parametri del Random Forest: piccole imprese Srl	
Nord	81
4.39 Tuning dei parametri del Random Forest: piccole imprese Srl	
Centro	82
4.40 Tuning dei parametri del Random Forest: piccole imprese Srl	
Sud/Isole	82
4.41 Tuning dei parametri del Random Forest: piccole imprese So-	
cietà Cooperative Nord	83
4.42 Tuning dei parametri del Random Forest: piccole imprese So-	
cietà Cooperative Centro	83

4.43	Tuning dei parametri del Random Forest: piccole imprese Società Cooperative Sud/Isole	84
4.44	Tuning dei parametri del Random Forest: medie imprese Spa Nord	84
4.45	Tuning dei parametri del Random Forest: medie imprese Spa Centro	85
4.46	Tuning dei parametri del Random Forest: medie imprese Spa Sud/Isole	85
4.47	Tuning dei parametri del Random Forest: medie imprese Srl Nord	86
4.48	Tuning dei parametri del Random Forest: medie imprese Srl Centro	86
4.49	Tuning dei parametri del Random Forest: medie imprese Srl Sud/Isole	87
4.50	Tuning dei parametri del Random Forest: medie imprese Società Cooperative Nord	87
4.51	Tuning dei parametri del Random Forest: medie imprese Società Cooperative Centro	88
4.52	Tuning dei parametri del Random Forest: medie imprese Società Cooperative Sud/Isole	88
4.53	Tuning dei parametri del Random Forest: grandi imprese Spa Nord	89

4.54	Tuning dei parametri del Random Forest: grandi imprese Spa	
	Centro	89
4.55	Tuning dei parametri del Random Forest: grandi imprese Spa	
	Sud/Isole	90
4.56	Tuning dei parametri del Random Forest: grandi imprese Srl	
	Nord	90
4.57	Tuning dei parametri del Random Forest: grandi imprese Srl	
	Centro	91
4.58	Tuning dei parametri del Random Forest: grandi imprese Srl	
	Sud/Isole	91
4.59	Tuning dei parametri del Random Forest: grandi imprese So-	
	cietà Cooperative Nord	92
4.60	Tipologie di funzioni di Kernel	95
4.61	Parametri specifici delle funzioni di Kernel	95
4.62	Setting dei parametri delle funzioni di Kernel	96
4.63	Tuning dei parametri del SVR: piccole imprese Spa Nord	96
4.64	Tuning dei parametri del SVR: piccole imprese Spa Centro . . .	97
4.65	Tuning dei parametri del SVR: piccole imprese Spa Sud/Isole . .	97
4.66	Tuning dei parametri del SVR: piccole imprese Srl Nord	98
4.67	Tuning dei parametri del SVR: piccole imprese Srl Centro	98
4.68	Tuning dei parametri del SVR: piccole imprese Srl Sud/Isole . .	99

4.69	Tuning dei parametri del SVR: piccole imprese Società Cooperative Nord	99
4.70	Tuning dei parametri del SVR: piccole imprese Società Cooperative Centro	100
4.71	Tuning dei parametri del SVR: piccole imprese Società Cooperative Sud/Isole	100
4.72	Tuning dei parametri del SVR: medie imprese Spa Nord	101
4.73	Tuning dei parametri del SVR: medie imprese Spa Centro . . .	101
4.74	Tuning dei parametri del SVR: medie imprese Spa Sud/Isole .	102
4.75	Tuning dei parametri del SVR: medie imprese Srl Nord	102
4.76	Tuning dei parametri del SVR: medie imprese Srl Centro . . .	103
4.77	Tuning dei parametri del SVR: medie imprese Srl Sud/Isole . .	103
4.78	Tuning dei parametri del SVR: medie imprese Società Cooperative Nord	104
4.79	Tuning dei parametri del SVR: medie imprese Società Cooperative Centro	104
4.80	Tuning dei parametri del SVR: medie imprese Società Cooperative Sud/Isole	105
4.81	Tuning dei parametri del SVR: grandi imprese Spa Nord	105
4.82	Tuning dei parametri del SVR: grandi imprese Spa Centro . . .	106
4.83	Tuning dei parametri del SVR: grandi imprese Spa Sud/Isole .	106
4.84	Tuning dei parametri del SVR: grandi imprese Srl Nord	107

4.85	Tuning dei parametri del SVR: grandi imprese Srl Centro . . .	107
4.86	Tuning dei parametri del SVR: grandi imprese Srl Sud/Isole . .	108
4.87	Tuning dei parametri del SVR: grandi imprese Società Coope- rative Nord	108
5.1	Risultati finali Piccole Imprese	110
5.2	Risultati finali Medie Imprese	111
5.3	Risultati finali Grandi Imprese	112

Elenco delle figure

2.1	Funzione di densità Ebitda Margin	27
2.2	Funzione di densità Valore della Produzione	28
2.3	Funzione di densità Costi della Produzione	28
2.4	Funzione di densità Totale Immobilizzazioni	29
2.5	Funzione di densità Totale Rimanenze	29
2.6	Funzione di densità Crediti a Breve	30
2.7	Funzione di densità Crediti Oltre	30
2.8	Funzione di densità Totale Disponibilità Liquide	31
2.9	Funzione di densità Debiti a Breve	31
2.10	Funzione di densità Debiti Oltre	32
2.11	Distribuzione di frequenza Divisioni territoriali	34
2.12	Distribuzione di frequenza Forme Giuridiche	34
2.13	Distribuzione di frequenza Microclassi	35
2.14	Funzione di densità Ebitda Margin Standardizzata	40

3.1	Clusterizzazione con DBSCAN	48
3.2	Clusterizzazione con K-MEANS	51
3.3	Clusterizzazione con K-MEANS	52
3.4	Clusterizzazione con K-MEANS	53
3.5	Clusterizzazione con K-MEANS.2	54
5.1	Range Ebitda Margin	113

Introduzione

La redditività aziendale è spesso espressa attraverso l'indice *Ebitda Margin*, derivante da alcune voci presenti all'interno del bilancio aziendale. L'*Ebitda Margin* è un'indice di redditività d'impresa, indicativo di quanto effettivamente questa sia in grado di generare reddito dalla gestione operativa. Lo scopo della seguente tesi è quello di analizzare imprese manifatturiere italiane appartenenti al settore alimentare. I dati utilizzati provengono dalla banca dati AIDA. In particolare, a partire da dati di bilancio degli anni 2014-2019, si è cercato di costruire un modello capace di prevedere i valori dell'*Ebitda Margin*, utilizzando dati di bilancio degli anni precedenti a quello di riferimento.

L'analisi è stata effettuata attraverso l'utilizzo di tecniche di *Machine Learning*. Nello specifico, la trattazione della tesi si struttura all'interno dei vari capitoli nel seguente modo:

il capitolo 1 è dedicato al *Business Understanding* e *Data Understanding*,

offrendo un'iniziale panoramica del contesto di riferimento, nonché delle caratteristiche e della semantica dei dati.

Il secondo capitolo tratta della *Cluster Analysis*, tecnica per suddividere le istanze del dataset in gruppi che siano il più omogenei possibili al loro interno e dissimilari tra loro. La ripartizione delle imprese è stata in primo luogo svolta su base dimensionale, per poi essere ulteriormente effettuata in relazione alla forma giuridica ed al territorio di appartenenza della sede legale di ogni azienda.

Nel capitolo successivo si procede con l'individuazione delle features più rilevanti ai fini dell'analisi, attraverso tecniche di *features selection* ed ulteriori studi volti all'individuazione di combinazioni di variabili che meglio performino nell'uso di differenti algoritmi per la regressione.

L'individuazione del migliore algoritmo tra quelli proposti per ogni casistica, rappresenta l'ultima fase del processo, attraverso un confronto della performance ottenuta in test.

Capitolo 1

2.1 Fonte del dataset

La seguente analisi è basata su dati provenienti dalla banca dati online AIDA, prodotta da Bureau van Dijk, multinazionale leader nella distribuzione di informazioni economico finanziarie. Nello specifico, AIDA è un database contenente informazioni economico finanziarie, dati anagrafici e merceologici di tutte le società di capitale italiane attive e fallite (ad esclusione di Banche, Assicurazioni ed Enti pubblici). Per ciascuna azienda AIDA offre:

- Bilancio dettagliato secondo lo schema della IV Direttiva CEE;
- indicatori economico finanziari precalcolati;
- dati indicizzati con uno storico fino a 10 anni;
- settore di attività (Ateco, SIC, NACE);

- informazioni anagrafiche complete;
- numero di dipendenti;
- valore delle azioni per società quotate;
- azionariato e partecipazioni;
- accesso alle note integrative – bilancio ottico;
- accesso alle operazioni di M&A;
- modulo di probabilità di default.

Il dataset contiene le variabili di bilancio dall'anno 2014 all'anno 2019 di imprese italiane. In particolare, sono presenti solo aziende manifatturiere di tre diverse forme giuridiche:

- Società per azioni (S.P.A.)
- Società a responsabilità limitata (S.R.L.)
- Società cooperative.

2.2 Business understanding

2.2.1 Analisi dell'industria alimentare italiana

Tra i diversi settori di attività disponibili, si è scelto di focalizzarsi sull'industria alimentare, corrispondente alla macro-classe 10, perché particolarmente caratteristica del panorama manifatturiero italiano e, di conseguenza, conte-

nente all'interno un numero di imprese tale da permettere un'analisi approfondita. Quest'ultima è composta all'interno da differenti micro-classi che rappresentano, nel dettaglio, i principali settori operativi da cui l'industria alimentare italiana è composta. La tabella 2.1 offre una panoramica sulla composizione della macro-classe di riferimento, nonché il numero di imprese appartenenti a ciascuna micro-classe:

MICROCLASSE	IMPRESE	% SUL TOT	MEDIA RICAVI	MEDIA EBITDAMARGIN
100	51	1.6%	27850.05	0.07
101	665	21.1%	30640.85	0.05
102	100	3.2%	19948.35	0.06
103	407	12.9%	22225.19	0.07
104	125	4.0%	51432.08	0.06
105	551	17.5%	25313.61	0.06
106	200	6.3%	29686.90	0.07
107	409	13.0%	26719.36	0.10
108	473	15.0%	30526.77	0.14
109	174	5.5%	26310.32	0.07

Tabella 2.1: Caratteristiche delle microimprese

MICROCLASSE	NORD	CENTRO	SUD	SPA	SRL	S. COOPERATIVE
100	51%	12%	37%	27%	73%	0%
101	73%	12%	15%	22%	74%	4%
102	37%	18%	45%	21%	77%	2%
103	36%	12%	52%	24%	66%	10%
104	38%	23%	39%	34%	62%	4%
105	63%	7%	30%	14%	54%	32%
106	68%	14%	19%	31%	66%	4%
107	64%	12%	24%	26%	73%	1%
108	67%	14%	19%	32%	67%	1%
109	67%	14%	19%	27%	70%	3%

Tabella 2.2: Caratteristiche delle microimprese.2

L'industria alimentare si presenta quindi molto composita al suo interno, con comparti le cui dimensioni medie variano in relazione all'area di appartenenza e alla forma giuridica di riferimento.

Un'analisi dettagliata del settore alimentare non può prescindere delle realtà territoriali all'interno delle quali le singole imprese operano. Un approccio territoriale nello studio di settore trova la sua *ratio* nell'imprescindibile legame presente tra industria e territorio. L'impresa è infatti considerata non come identità a sé stante ma, al contrario, in continuo rapporto con l'ambiente di appartenenza.

In particolare si nota come, nella maggior parte delle microclassi, le imprese si concentrino al Nord, con una prevalenza di SRL sulle restanti forme giuridiche.

2.3 Data understanding

Tra le diverse voci di bilancio presenti nello Stato Patrimoniale e Conto economico di ciascuna impresa, solo alcune sono state ritenute come potenzialmente utili ai fini dell'analisi.

In particolare, le variabili selezionate dal Conto Economico sono:

- **VALORE DELLA PRODUZIONE:** incremento di valore che si verifica nell'ambito della produzione e distribuzione di beni e servizi finali, grazie ai fattori produttivi adoperati (capitale e lavoro) a partire da beni e risorse primarie iniziali.
- **COSTI DELLA PRODUZIONE:** la somma di tutti i costi di utilizzo dei fattori impiegati nel processo produttivo.

Le voci di bilancio dello Stato Patrimoniale sono:

ATTIVO

- **TOTALE IMMOBILIZZAZIONI:** voce di bilancio che deriva dalla somma di:
 - Immobilizzazioni materiali: beni aziendali ad uso durevole e pluriennale quali terreni, fabbricati, impianti, macchinari, attrezzature

- industriali e commerciali, immobilizzazioni in corso e acconti.
- Immobilizzazioni immateriali: beni aziendali non tangibili ad uso durevole e pluriennale quali diritti di brevetto industriale e diritti di utilizzazione delle opere di ingegno, concessioni, licenze, marchi, avviamento, immobilizzazioni in corso e acconti.
 - Immobilizzazioni finanziarie: titoli, partecipazioni e crediti che l'azienda detiene a titolo di investimento durevole e non a fine speculativo di breve periodo.
- TOTALE RIMANENZE: rimanenze di magazzino che includono materie prime, sussidiarie e di consumo, prodotti in corso di lavorazione e semilavorati, lavori in corso su ordinazione, prodotti finiti e merci, acconti.
 - TOTALE DISPONIBILITÀ LIQUIDE: depositi bancari e postali, assegni bancari e circolari, denaro e valori in cassa.
 - CREDITI A BREVE: crediti verso clienti esigibili entro l'esercizio successivo.
 - CREDITI OLTRE L'ANNO: crediti verso clienti esigibili oltre l'esercizio successivo.

PASSIVO

- DEBITI A BREVE: debiti verso fornitori esigibili entro l'esercizio successivo.
- DEBITI OLTRE L'ANNO: debiti verso fornitori esigibili oltre l'esercizio successivo.

In aggiunta, sebbene non presente tra le voci di bilancio, è stata aggiunta la variabile DIPENDENTI, che mostra il numero di dipendenti all'interno di ciascuna impresa.

La variabile dipendente, l'EBITDA MARGIN, è un indice che deriva dall'Ebitda. Quest'ultimo, acronimo inglese di *Earnings before Interest, Taxes, Depreciation and Amortization*, è un indicatore di redditività utilizzato nell'analisi di bilancio per la valutazione di un'azienda. Essendo una grandezza in grado di esprimere una misura di reddito al lordo degli ammortamenti, risulta fondamentale per valutare quanto un'impresa sia in grado di generare profitti derivanti dalla gestione ordinaria, ovvero l'insieme di operazioni normalmente svolte dall'azienda nello svolgimento dell'attività per cui è nata.

Per il calcolo dell'Ebitda, uno degli schemi maggiormente utilizzati è quello a valore aggiunto. Si tratta di misurare l'entità del valore che l'azienda, attraverso il processo di trasformazione posto in essere dal ciclo operativo, è in grado di aggiungere alle materie prime e agli altri fattori della produzione

acquisiti all'esterno.

Di seguito la formula matematica da cui si ottiene l'Ebitda:

$$Ebitda = ValoreDellaProduzione - CostidellaProduzione \quad (2.1)$$

Ai fini di una valutazione più accurata di quanto effettivamente l'impresa sia in grado di generare reddito dalla gestione operativa, si è soliti far riferimento all'Ebitda Margin. L'Ebitda Margin è un indicatore derivante dal rapporto tra l'Ebitda e il valore della produzione.

$$EbitdaMargin = \frac{Ebitda}{ValoreDellaProduzione} \quad (2.2)$$

È un indice di redditività che misura il risultato lordo della gestione operativa (Ebitda) sui ricavi di vendita. Più nello specifico, potrebbe anche essere interpretato come un indice di situazione finanziaria, in quanto ha l'ulteriore significato di autofinanziamento della gestione operativa. In tale ottica, l'indice segnala la misura percentuale di produzione di mezzi finanziari interni da parte dei ricavi di vendita.

Sebbene a valori più elevati dell'Ebitda Margin, corrisponda una migliore situazione aziendale, l'indice non ha parametri logici di riferimento. Esso dovrebbe essere confrontato con valori medi di settore. Ad esempio, per il settore manifatturiero, una soglia di adeguatezza è fissata circa all'8%.

2.3.1 Struttura del dataset

il dataset originale è composto da 14 variabili di diversa natura, come riportato nella tabella 2.3:

Tabella 2.3: Struttura del dataset

VARIABILE	TIPOLOGIA
VALORE DELLA PRODUZIONE	continua
COSTI DELLA PRODUZIONE	continua
TOTALE IMMOBILIZZAZIONI	continua
TOTALE RIMANENZE	continua
CREDITI A BREVE	continua
CREDITI OLTRE	continua
TOTALE DISPONIBILITÀ LIQUIDE	continua
DEBITI A BREVE	continua
DEBITI OLTRE	continua
DIPENDENTI	continua
EBITDA MARGIN	continua
AREA	categorica
FORMA GIURIDICA	categorica
MICROCLASSE	categorica

2.3.2 Variabili continue

Statistiche base

Con riferimento alle sole variabili continue, il dataset originale ne contiene i valori di ciascun anno dal 2014 al 2019. Al solo fine di fornire un rappresentazione chiara della composizione del dataset, si è optato per calcolare la media tra anni di ciascuna variabile, in modo da ottenere una singola colonna per ciascuna voce. La tabella 2.4 offre una statistica descrittiva delle variabili ottenute:

Variable	Mean	SD	Skewness	Kurtosis
VALORE DELLA PRODUZIONE	28239.08	91700.27	15.59	351.84
COSTI DELLA PRODUZIONE	27789.92	90220.56	15.45	346.29
TOTALE IMMOBILIZZAZIONI	11341.69	80790.08	27.74	928.61
TOTALE RIMANENZE	4197.37	12020.79	8.83	109.31
CREDITI A BREVE	6968.27	18932.75	12.67	224.95
CREDITI OLTRE	218.46	1416.48	28.55	1121.18
TOTALE DISPONIBILITÀ LIQUIDE	1804.42	8775.19	23.71	715.02
DEBITI A BREVE	10435.65	29656.87	11.62	211.60
DEBITI OLTRE	3222.30	12458.34	13.64	249.03
DIPENDENTI	58.62	207.33	18.23	462.76
EBITDA MARGIN	0.07	0.76	38.12	2186.95

Tabella 2.4: Distribuzioni delle variabili in migliaia di euro

Come la tabella suggerisce, la presenza di un'elevata deviazione standard in tutte le variabili indica una significativa variabilità all'interno del campione di riferimento. Ciò potrebbe essere causato dalla presenza di imprese di differenti dimensioni e collocazione territoriale.

Le due misure *Skewness* e *Kurtosis* descrivono quanto la distribuzione della variabile è simmetrica o meno. Risultano particolarmente significative nell'indicare dati non distribuiti normalmente.

Come si nota dalla tabella sopra riportata, tutte le variabili presentano valori positivi della *Skewness*, ad indicare che la maggior parte dei valori si concentra sul lato sinistro del grafico. In questo caso, la mediana (l'unità che occupa la posizione centrale nella distribuzione ordinata dei valori) è sempre più piccola della media aritmetica perché, a differenza di quest'ultima, non è influenzata dalla presenza di dati anomali.

La maggiorparte delle osservazioni si concentrano, quindi, in un intervallo che corrisponde a valori piccoli delle singole variabili. Con riferimento al dominio applicativo, valori bassi di voci di bilancio potrebbero indicare la presenza di una moltitudine di imprese di piccola o media dimensione, fenomeno che rispecchia a pieno il panorama imprenditoriale italiano.

Analogamente alla *Skewness*, anche la *Kurtosis* assume valori positivi in tutte le variabili. Un eccesso di *Kurtosis* positiva suggerisce la presenza di una distribuzione leptocurtica, dove il numero valori anomali presenti è di gran

lunga superiore a quello di una distribuzione normale.

Nel caso esaminato risulta evidente un numero relativamente basso di istanze con elevati valori di ciascuna variabile, valori considerabili anomali rispetto alla maggiorparte dei dati. Potrebbe trattarsi, infatti, della presenza di poche grandi imprese che operano nel settore alimentare.

Distribuzioni univariate

I grafici riportati in Figura 2.1-2.10 offrono una visualizzazione delle distribuzioni delle variabili, mostrando le asimmetrie presenti in ognuna di esse.

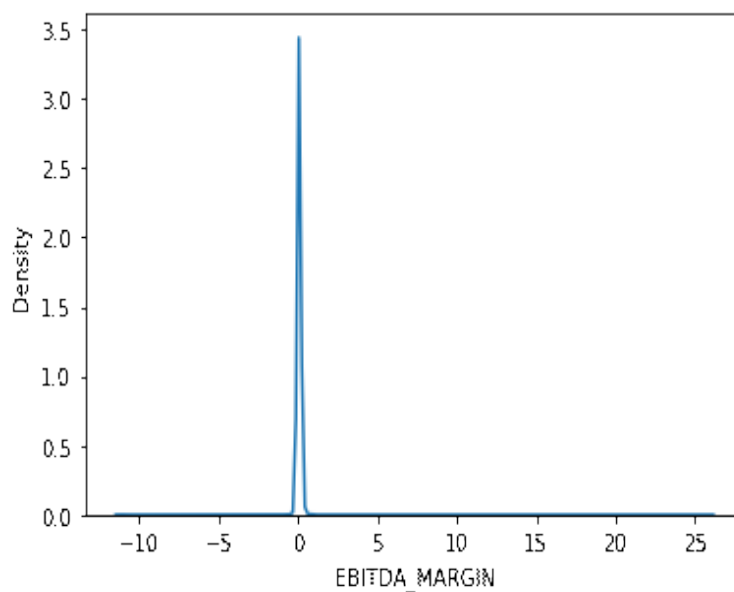


Figura 2.1: Funzione di densità Ebitda Margin

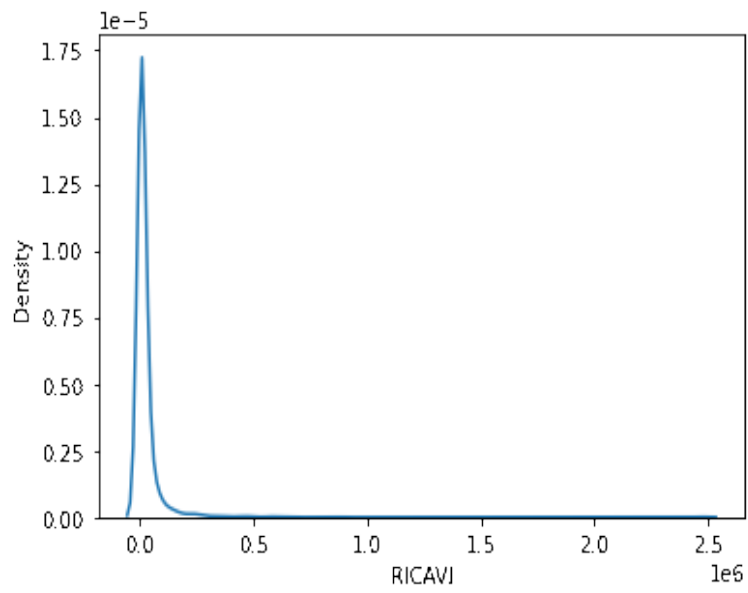


Figura 2.2: Funzione di densità Valore della Produzione

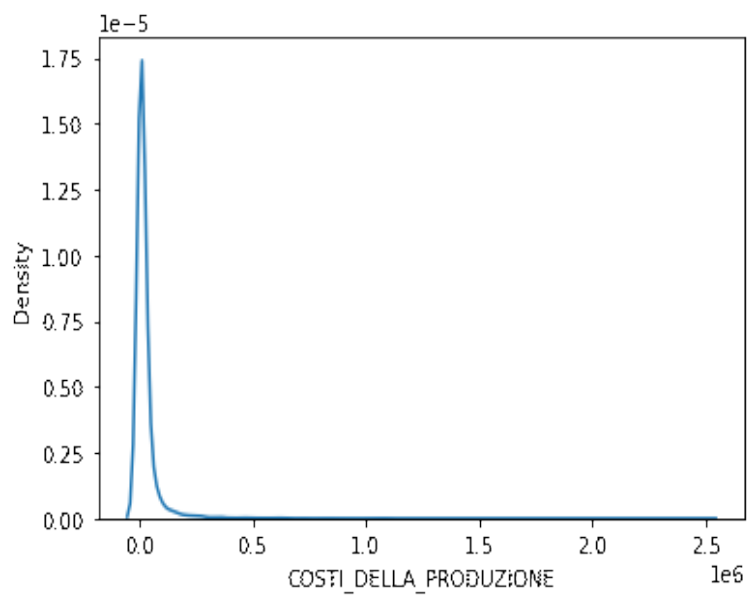


Figura 2.3: Funzione di densità Costi della Produzione

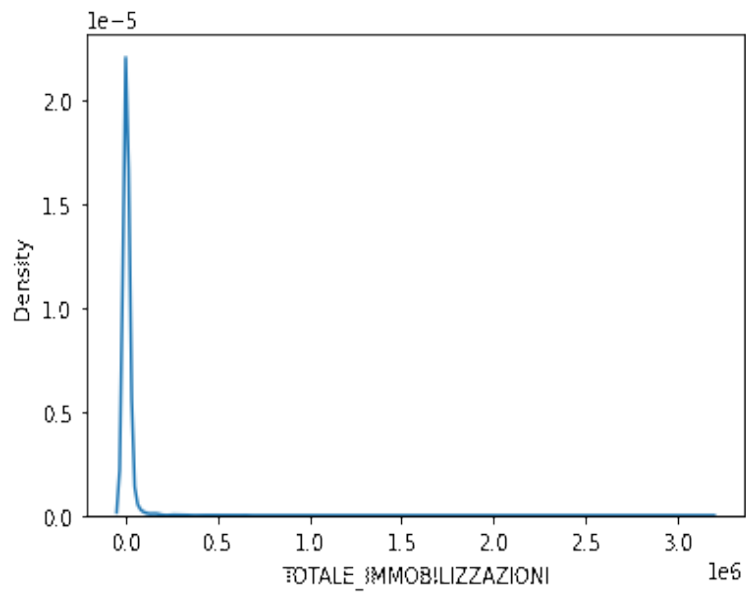


Figura 2.4: Funzione di densità Totale Immobilizzazioni

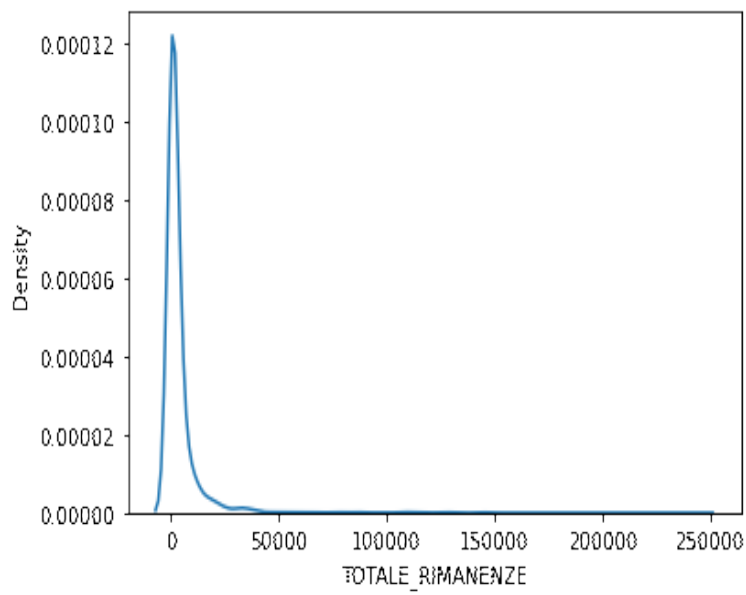


Figura 2.5: Funzione di densità Totale Rimanenze

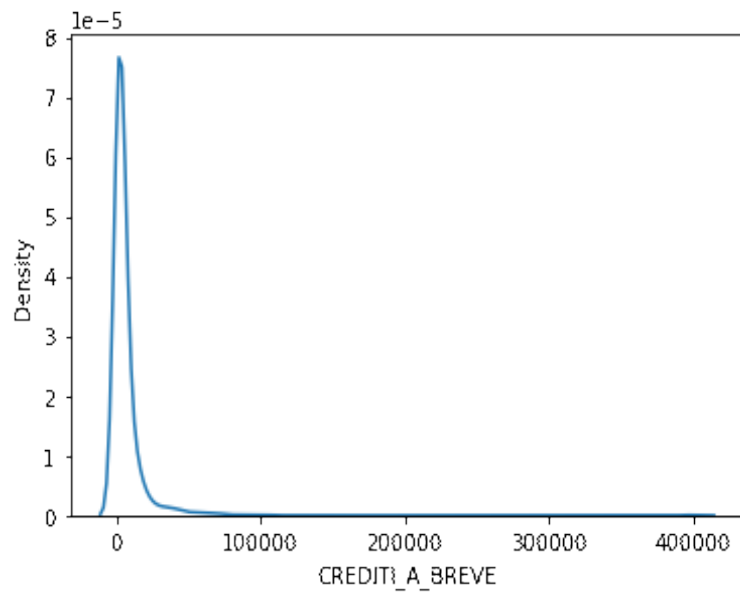


Figura 2.6: Funzione di densità Crediti a Breve

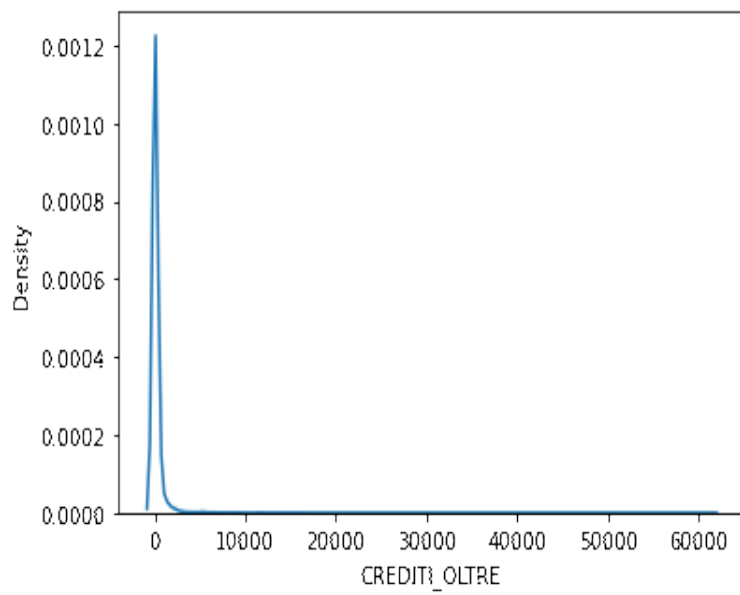


Figura 2.7: Funzione di densità Crediti Oltre

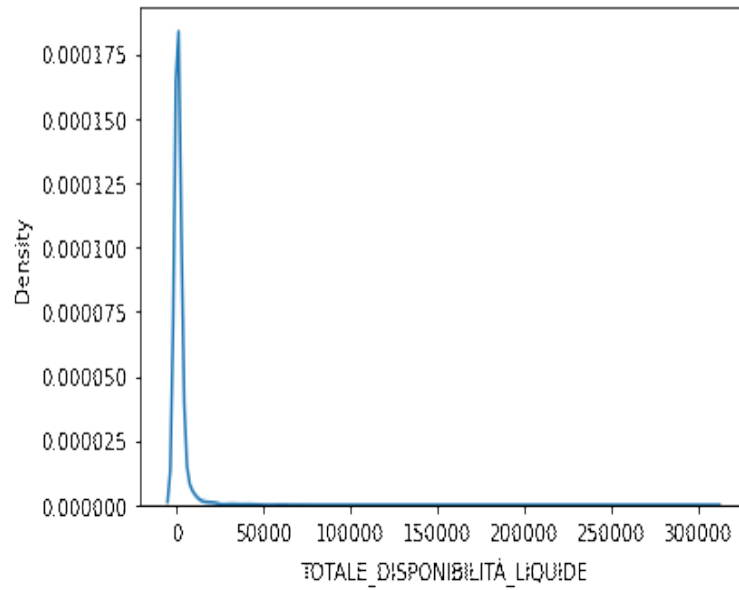


Figura 2.8: Funzione di densità Totale Disponibilità Liquide

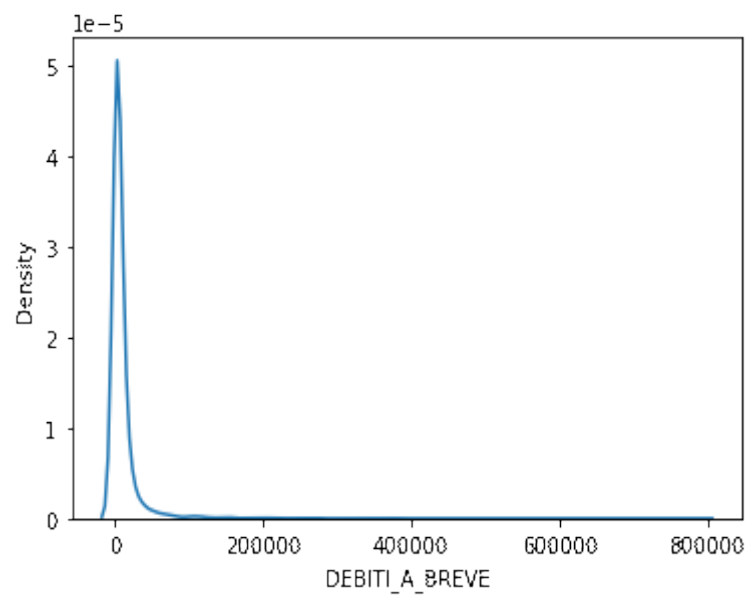


Figura 2.9: Funzione di densità Debiti a Breve

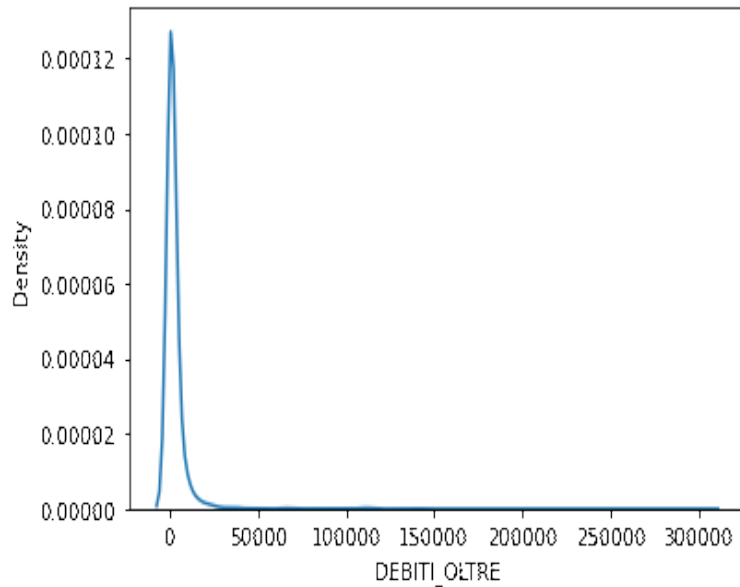


Figura 2.10: Funzione di densità Debiti Oltre

2.3.3 Variabili categoriche

Le restanti variabili sono variabili categoriche:

- AREA: può assumere valori Nord, Centro, Sud, Isole
- FORMA GIURIDICA: può assumere valori Spa, Srl, Società Cooperativa
- MICROCLASSE: può assumere valori
 - 100: nessuna particolare definizione
 - 101: lavorazione e conservazione di carne e produzione di prodotti a base di carne

- 102: lavorazione e conservazione di pesce, crostacei e molluschi
- 103: lavorazione e conservazione di frutta e ortaggi
- 104: produzione di oli e grassi vegetali e animali
- 105: industria lattiero-casearia
- 106: lavorazione delle granaglie, produzione di amidi e di prodotti amidacei
- 107: produzione di prodotti da forno e farinacei
- 108: produzione di altri prodotti alimentari
- 109: produzione di prodotti per l'alimentazione degli animali

Distribuzioni univariate

I grafici riportati in Figura 2.11-2.13 offrono una visualizzazione delle distribuzioni delle variabili categoriche.

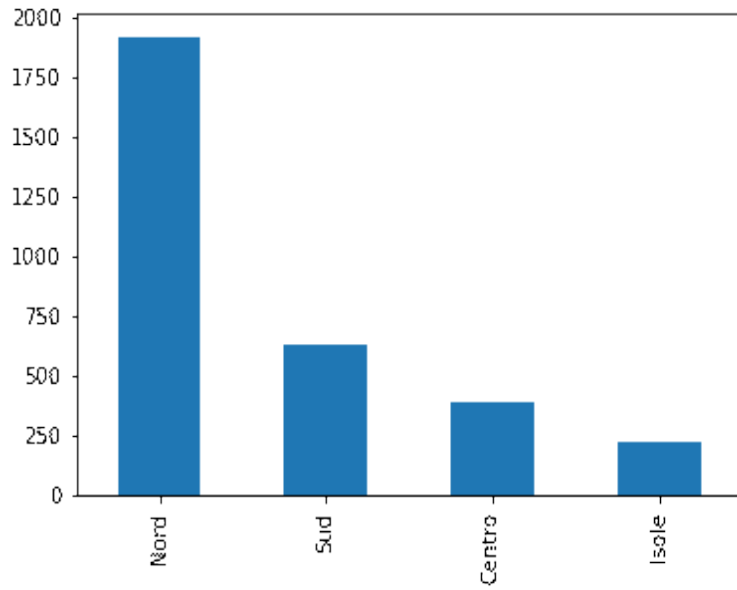


Figura 2.11: Distribuzione di frequenza Divisioni territoriali

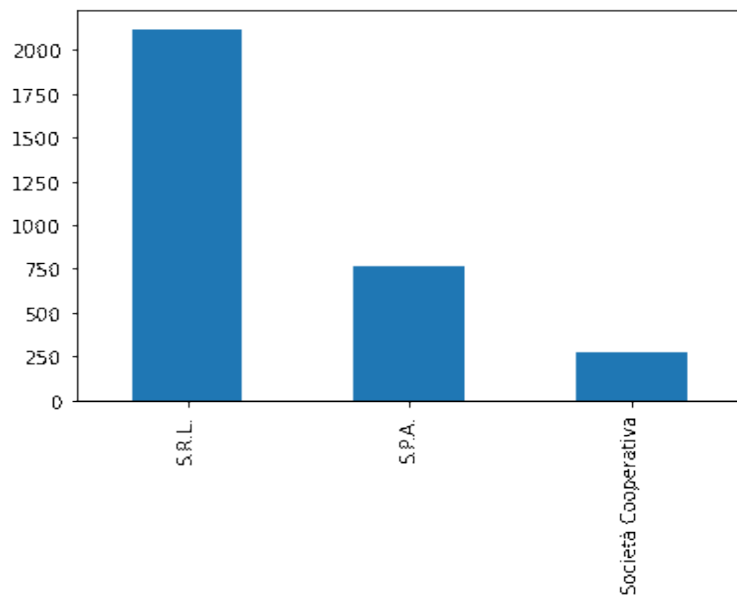


Figura 2.12: Distribuzione di frequenza Forme Giuridiche

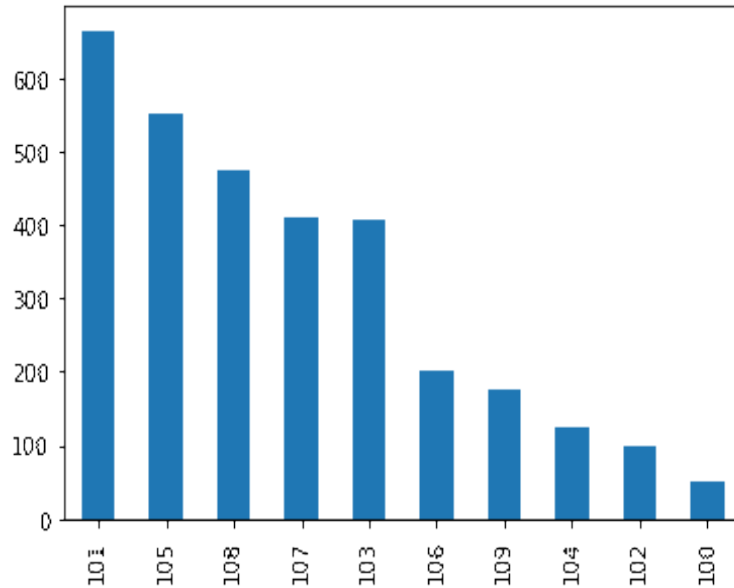


Figura 2.13: Distribuzione di frequenza Microclassi

2.4 Preprocessing

2.4.1 Data Cleaning

Al fine di ottenere dati di maggior qualità possibile, si è proceduto con un iniziale trattamento degli ND, valori non disponibili nel dataset originale. La motivazione della mancanza di alcuni dati risiede nel fatto che non tutte le imprese, al momento del deposito del bilancio, hanno correttamente indicato i valori di tutte le voci di bilancio. Si è ritenuto opportuno non stimare gli ND assegnandogli il valore medio o mediano della relativa variabile di appartenenza. Trattandosi infatti di dati sociali, la stima sarebbe potuta essere non rappresentativa del fenomeno. Per questa ragione si è optato per rimuovere

le istanze in cui era presente almeno un valore ND, riducendo il numero delle istanze da 3155 a 2806, con una perdita non significativa di dati.

2.4.2 Data Integration

Nel normale svolgimento dell'attività d'impresa potrebbero risultare d'impatto anche alcuni fenomeni esterni all'impresa stessa. In particolare, dinamiche di settore e variabili macroeconomiche potrebbero incidere sulla redditività d'impresa.

Per questo motivo sono state identificate e aggiunte al dataset originale alcune variabili esterne, ritenute potenzialmente utili ai fini dell'analisi. Analogamente alle voci di bilancio, anche le nuove features riportano tutti i valori dal 2014 al 2019. Esse si sostanziano essere:

- **PIL PRO CAPITE**: indicatore che esprime il livello di ricchezza di un determinato territorio. Deriva dal rapporto tra il Prodotto Interno Lordo (PIL) e il numero di abitanti del territorio preso in considerazione.
- **INDICE DEI PREZZI AL CONSUMO**: misura statistica formata dalla media dei prezzi ponderati per mezzo di uno specifico paniere di beni e servizi. Tale paniere ha come riferimento le abitudini di acquisto di un consumatore medio.
- **GRADO DI APERTURA COMMERCIALE**: Indice dato dal rapporto tra la somma delle importazioni ed esportazioni ed il Pil. Viene comu-

nemente utilizzato per per misurare il livello di integrazione economico-commerciale internazionale di un paese o di un'area geoeconomica. In questo caso l'indice fa riferimento al settore alimentare italiano.

- **BUSINESS CONFIDENCE INDEX:** per le aziende e il mercato azionario, la fiducia delle imprese indica le aspettative delle imprese, sulla base di indagini su produzione, ordini e prodotti finiti nel settore.
- **CONFIDENCE CONFIDENCE INDEX:** indice che misura il livello di fiducia dei consumatori nei confronti della loro situazione finanziaria prevista. L'indice si basa sul presupposto che se i consumatori sono ottimisti, spenderanno di più e stimoleranno l'economia, ma se sono pessimisti, i loro modelli di spesa potrebbero portare a un rallentamento economico.
- **INFLAZIONE:** aumento progressivo del livello medio generale dei prezzi, che determina una diminuzione progressiva del potere di acquisto (cioè del valore) della moneta. Il fenomeno può avere molteplici cause, sia reali sia monetarie.
- **ISIC:** indice sintetico di competitività dell'agroalimentare ossia un indicatore statistico che fornisce una misura delle performance competitive dell'agricoltura e dell'industria alimentare a livello regionale. Permette di misurare il posizionamento del sistema agricolo e alimentare delle diverse regioni, e la sua evoluzione nel tempo, rispetto alla media

nazionale.

L'integrazione delle nuove variabili, dove possibile, è stata effettuata basandosi sulla corrispondenza dei valori in base al territorio di riferimento. In particolare, i valori di Grado di Apertura Commerciale e ISIC corrispondono per tutti gli anni al territorio di appartenenza dell'impresa. Le restanti variabili, invece, differiscono solo negli anni.

La tabella 2.5 offre una statistica descrittiva delle nuove features introdotte:

Variabile	Media	SD	Range
PIL PRO CAPITE	32832.18	0.00	[32832.18, 32832.18]
INDICE PREZZI AL CONSUMO	103.53	0.00	[103.53, 103.53]
GRADO APERTURA COMMERCIALE	2.45	0.71	[1.08, 2.97]
BUSINESS CONFIDENCE INDEX	100.78	0.00	[100.78, 100.78]
CONSUMER CONFIDENCE INDEX	100.22	0.00	[100.22, 100.22]
INFLAZIONE	101.00	0.00	[101.00, 101.00]
ISIC	104.38	1.00	[100.93, 104.88]

Tabella 2.5: Distribuzione delle nuove variabili macroeconomiche

Analogamente a come proceduto con le voci di bilancio, anche i valori delle nuovi variabili sono stati mediati tra anni, al solo fine di offrire una panoramica chiara del comportamento delle features.

2.4.3 Data Normalization

La standardizzazione è la fase del preprocessing che permette di rendere le scale di variabili, che hanno media e deviazione standard diverse tra loro, il più simili possibili, permettendo quindi di confrontarle.

Nel caso esaminato, la tipologia di standardizzazione adottata è stata la *Z-standardization*. I valori ottenuti derivano dal seguente calcolo:

$$Z = \frac{x - \mu}{\sigma} \quad (2.3)$$

Per ottenere il valore standardizzato, al valore originale si sottrae la media e il risultato è diviso per la deviazione standard della variabile di riferimento.

Con particolare riferimento alla variabile dipendente, l'*Ebitda Margin*, il nuovo range risulta essere (-31.32, 13.62) Il grafico in Figura 2.14 mostra la distribuzione della variabile standardizzata. La maggiorparte dei valori si concentra nel più ristretto range (-3, 4), con poche istanze che assumono valori significativamente lontani dalla mediana.

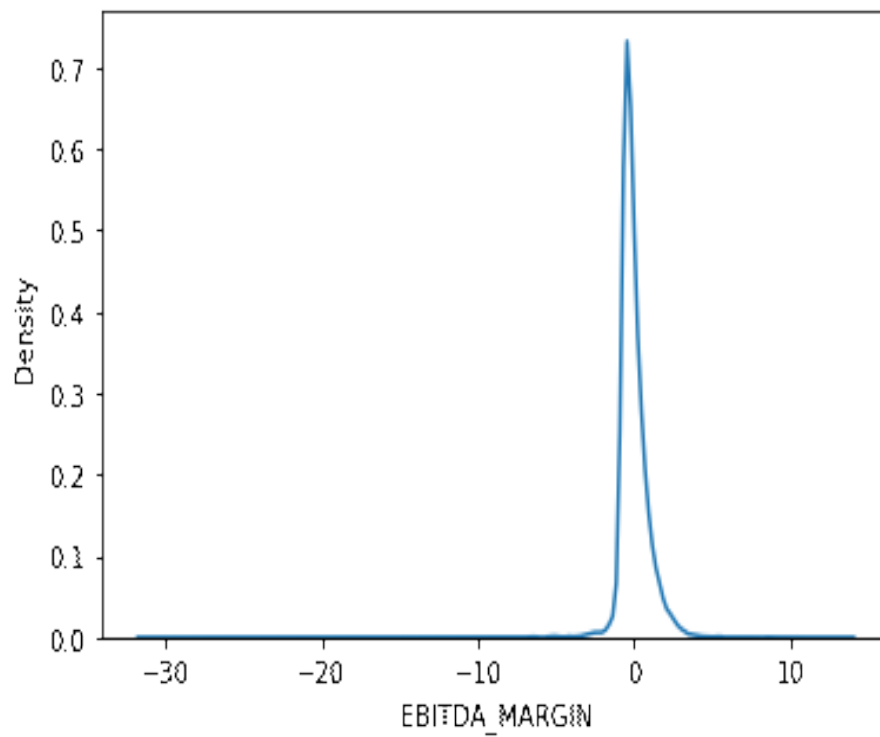


Figura 2.14: Funzione di densità Ebidta Margin Standardizzata

La Tabella 2.6 mostra i nuovi range di tutte le variabili del dataset:

Variabile	Min Range	Max Range
VALORE DELLA PRODUZIONE	-0.305	32.885
COSTI DELLA PRODUZIONE	-0.305	33.256
TOTALE IMMOBILIZZAZIONI	-0.137	39.795
TOTALE RIMANENZE	-0.346	21.624
CREDITI A BREVE	-0.377	24.615
CREDITI OLTRE	-0.185	36.062
TOTALE DISPONIBILITÀ LIQUIDE	-0.16	59.477
DEBITI A BREVE	-0.357	37.712
DEBITI OLTRE	-0.245	46.111
DIPENDENTI	-0.289	34.933
PIL PRO CAPITE	-1.184	1.436
INDICE PREZZI AL CONSUMO	-1.18	1.205
GRADO APERTURA COMMERCIALE	-2.053	0.984
BUSINESS CONFIDENCE INDEX	-1.41	1.366
CONSUMER CONFIDENCE INDEX	-1.746	1.374
INFLAZIONE	-0.801	1.744
ISIC	-1.596	1.437
EBITDA MARGIN	-31.321	13.619

Tabella 2.6: Range delle variabili standardizzate

Capitolo 2

3.1 Cluster

Il seguente capitolo tratta di differenti tecniche di clusterizzazione delle imprese presenti all'interno del dataset. La *Cluster analysis* è una tecnica descrittiva non supervisionata di Data Mining che ha come obiettivo principale quello di dividere i dati in gruppi che siano al loro interno il più coesi e omogenei possibili, con istanze dalle caratteristiche simili. Essendo una tecnica non supervisionata, il cluster cerca di catturare la struttura naturale dei dati originali senza fare uso di nessun dato esterno.

I cluster devono essere ottenuti in modo tale da minimizzare la distanza intra-gruppo e massimizzare, invece, quella inter-gruppo, in modo che oggetti di cluster diversi differiscano il più possibile.

3.2 DBSCAN

Dalla descrizione delle singole variabili di bilancio precedentemente trattata, è emersa la presenza di istanze *outliers*, che presentano cioè valori significativamente anomali rispetto alla maggiorparte dei dati. Essendo lontani dalla mediana, la presenza di outliers potrebbe compromettere una successiva analisi. Al riguardo, si è convenuto procedere con l'utilizzo della tecnica di clusterizzazione del DBSCAN, per l'identificazione e l'eventuale rimozione degli stessi.

3.2.1 Definizione

Il DBSCAN è una tecnica di clusterizzazione con un approccio partitivo esclusivo, che prevede cioè la divisione del dataset in cluster dove la singola istanza appartiene solo ad un gruppo, senza creare sovrapposizione tra essi. È una tipologia *density-based*, basata cioè sulla densità. L'idea di fondo è quella di interpretare il cluster come una regione ad alta intensità, circondata da una a bassa densità.

Limite principale del DBSCAN è quello di non riuscire a fornire una descrizione generale delle caratteristiche delle istanze presenti all'interno di ciascun cluster. Per questo motivo, la tecnica è stata implementata in fase iniziale con lo scopo di identificare valori anomali all'interno del dataset.

3.2.2 Funzionamento

Il DBSCAN riconosce tre tipologie di istanze:

- CORE POINT: elemento del dataset nel cui vicinato (sfera il cui raggio corrisponde al parametro Eps) sono presenti almeno un numero di elementi pari ad un certo *threshold*. Tale soglia è definita dal parametro *Min Points*.
- BORDER POINT: oggetto che non è definibile come *core point*, ma cade nel vicinato di uno o più *core points*.
- NOISE POINT: tutti gli oggetti restanti. Sono considerabili come rumore e quindi da rimuovere o trattare separatamente. All'interno di questa definizione rientrano anche gli *outliers*, in quanto valori anomali.

Eps e *Min Points* sono i due parametri da settare per poter ottenere la distinzione nelle tre forme di oggetti. Nello specifico essi si sostanziano essere come:

- Eps : raggio che determina le dimensioni della sfera intorno ad un oggetto.
- *Min Points*: numero di elementi minimi presenti all'interno della sfera di raggio Eps , per far sì che un punto sia definibile come *core point*.

L'algoritmo DBSCAN è riportato in Algoritmo 1:

Algorithm 1 Algoritmo DBSCAN

- 1: scansione e definizione di ciascun punto come *core*, *border* o *noise point*.
 - 2: rimozione dei *noise points*.
 - 3: definire se la distanza tra due *core points* è minore di *Eps*.
 - 4: assegnazione di core point con distanza minore dell' *Eps* allo stesso cluster.
 - 5: assegnazione di ciascun *border point* al cluster del relativo *core point*.
-

3.2.3 Metriche

La metrica utilizzata nella seguente analisi per valutare la bontà della clusterizzazione con DBSCAN è stata quella della *Silhouette*. La *Silhouette* è una metrica non supervisionata che calcola la similarità di un punto con gli oggetti del cluster di appartenenza e la confronta con la dissimilarità dagli oggetti degli altri cluster.

Per ciascun punto è calcolata la similarità con il cluster di appartenenza:

$$a(x) = \frac{1}{|C_i| - 1} \sum_{y \in C_i, y \neq x} proximity(x, y) \quad (3.1)$$

e la dissimilarità con il cluster più vicino al proprio:

$$b(x) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{y \in C_k} proximity(x, y) \quad (3.2)$$

Si definisce quindi *Silhouette* di un punto:

$$s(x) = \begin{cases} 1 - a(x)/b(x) & \text{se } a(x) < b(x) \\ 0 & \text{se } a(x) = b(x) \\ b(x)/a(x) - 1 & \text{se } a(x) > b(x) \end{cases} \quad (3.3)$$

Per calcolare la *Silhouette* totale, si procede con una media delle singole. Valori medi prossimi a 1 indicano una buona clusterizzazione.

3.2.4 Analisi

La clusterizzazione delle istanze è stata effettuata prendendo a riferimento la variabile Valore della Produzione (che deriva dalla media di Valore della Produzione negli anni 2014-2019). Tale scelta è basata sul fatto che l'ammontare del valore della produzione di un'impresa potrebbe essere indicativo della dimensionalità della stessa, riuscendo quindi a discriminare imprese di grandezza differente.

Dal momento che sia il DBSCAN che il K-Means, successivamente descritto, si basano sul concetto di distanza, la metrica presa a riferimento è stata quella della *Squared Euclidean Distance*, espressa come:

$$d^2(p, q) = (p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2 \quad (3.4)$$

dove p e q rappresentano due punti nello spazio euclideo e n lo spazio dimensionale.

Al fine di individuare i valori ottimi dei parametri *Eps* e *Min Points*, si è optato per effettuare un loop, attraverso l'operatore di Rapid Miner *Loop Parameters*. Il loop permette di ottenere una serie di combinazioni dei parametri, all'interno del range impostato per ciascuno, a cui è associata il valore della metrica.

Il setting dei parametri è stato realizzato nel seguente modo:

Parametro	Min	Max	Steps	Scale
EPS	0.1	0.8	30	logarithmic
MIN POINTS	5	30	25	linear

Tabella 3.1: Setting dei parametri del DBSCAN

Lo scopo, infatti, è quello di selezionare la combinazione con il valore della metrica di riferimento più alto. La migliore combinazione risulta essere quella con *Eps*=0.115 e *Min Points*=24. In figura 3.1 sono riportati i risultati ottenuti usando tali parametri.

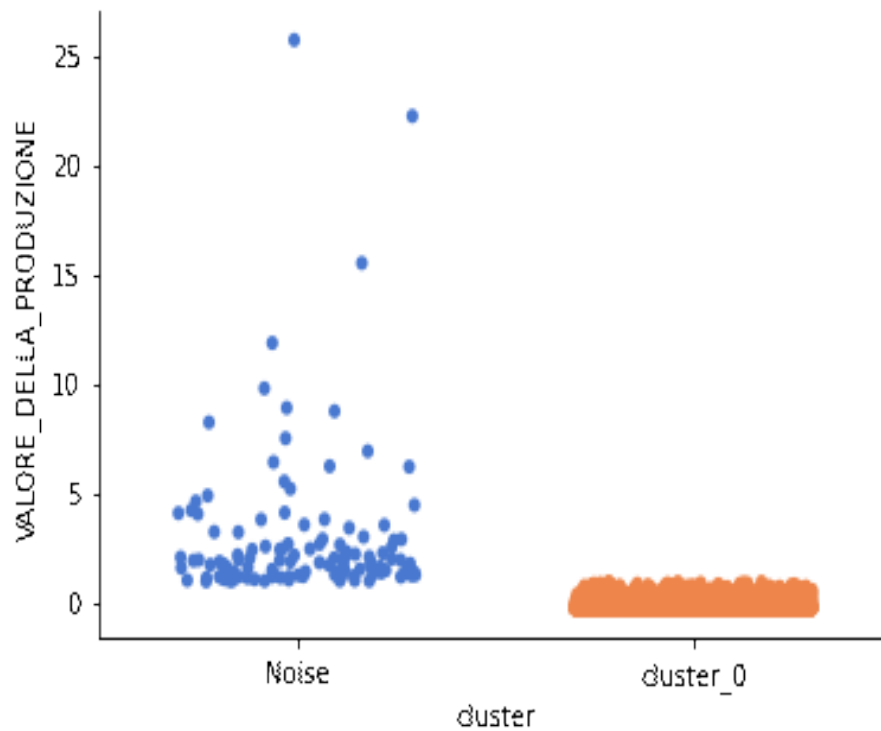


Figura 3.1: Clusterizzazione con DBSCAN

L'algoritmo discrimina il *Noise*, inteso come istanze con valori anomali e il restante cluster le cui istanze presentano valori normalizzati molto vicino allo zero. Il cluster *Noise* è formato da 119 imprese, mentre il cluster principale dalle restanti 2686.

I valori anomali individuati corrispondono a imprese con un Valore della Produzione molto elevati, ad indicare probabilmente aziende di elevatissima dimensionalità. Dal momento che quest'ultime sono *outliers*, verranno rimosse dall'analisi.

Il secondo cluster formatosi presenta invece un numero elevato di istanze, fittamente concentrate nello spazio. Per questo motivo si è proceduto con un'ulteriore clusterizzazione attraverso la tecnica del K-Means.

3.3 K-MEANS

Il K-Means si basa sugli stessi presupposti e condivide le medesime finalità del DBSCAN. A differenza di quest'ultimo, però, è un algoritmo di clusterizzazione *Prototype-based*, che presuppone cioè che ogni cluster non è rappresentato dall'insieme dei suoi punti, ma dal valore del proprio centroide.

3.3.1 Funzionamento

Per il K-Means l'unico parametro presente è il k , che rappresenta il numero di cluster da formare definito a priori.

L'algoritmo del K-Means è brevemente descritto in Algoritmo 2.

Algorithm 2 Algoritmo K-Means

- 1: selezione di K punti come centroidi iniziali.
 - 2: **repeat**
 - 3: formazione di K cluster assegnando ogni punto al centroide più vicino.
 - 4: ricalcolo del centroide di ogni cluster.
 - 5: **until** i centroidi non cambiano.
-

3.3.2 Metriche

La metrica utilizzata per valutare la bontà della clusterizzazione è il *David Bouldin score*, che calcola la similarità tra due gruppi.

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \quad (3.5)$$

dove:

S_i : dispersione intra-cluster i

S_j : dispersione intra-cluster j

M_{ij} : distanza tra centroidi del cluster i e j

In questo caso, minore è il valore della metrica, migliore è la clusterizzazione ottenuta.

3.3.3 Analisi

Analogamente a come proceduto con il DBSCAN, per individuare il valore ottimo del parametro k è stato effettuato un *Loop Parameters*. Il setting è è riportato in Tabella 3.2:

Parametro	Min	Max	Steps	Scale
K	2	12	10	linear

Tabella 3.2: Setting dei parametri del K-MEANS

La figura 3.2 mostra i risultati del loop, dove ad ogni possibile valore di K

corrisponde il valore della metrica *David Bouldin score*. Come si può osservare, il numero ottimale di è 2.

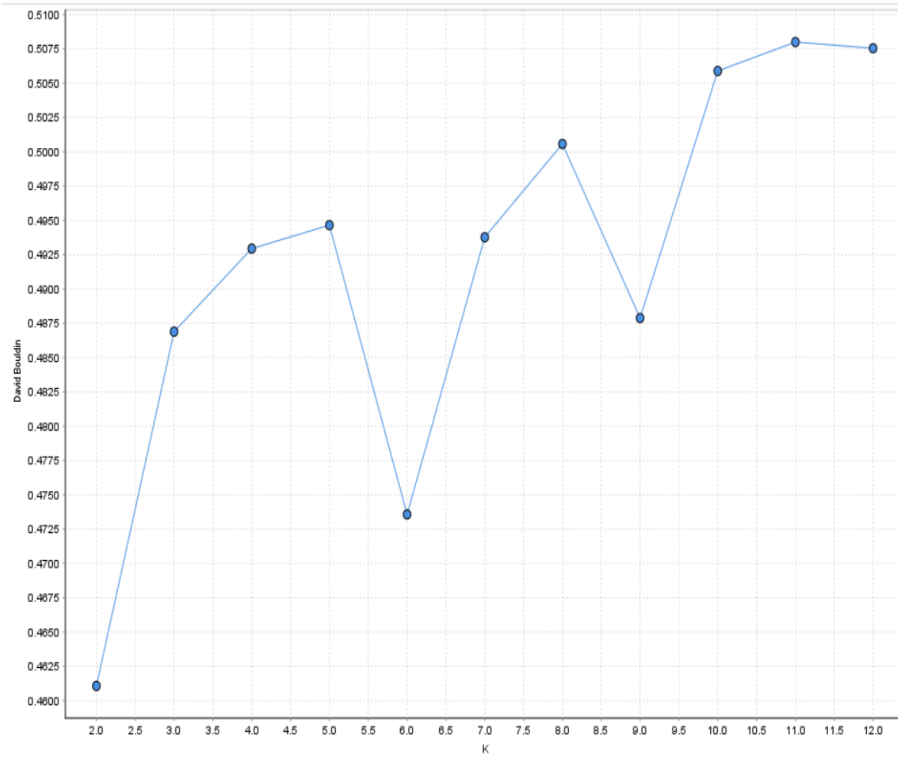


Figura 3.2: Clusterizzazione con K-MEANS

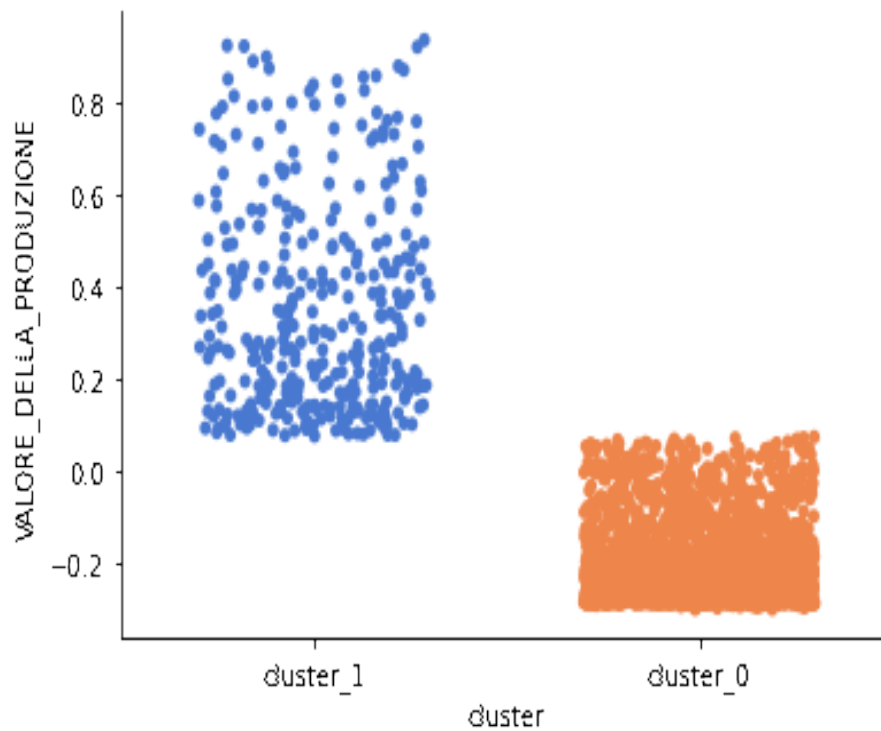


Figura 3.3: Clusterizzazione con K-MEANS

La tabella 3.3 descrive i singoli cluster:

Cluster	Numero Imprese	Centroide
cluster_0	2342	11078,24
cluster_1	344	65187,22

Tabella 3.3: Caratteristiche dei cluster

Il cluster 1 rappresenta istanze con Valore della Produzione maggiore di quelle contenute nel cluster 0. Esso rappresenta l'insieme delle imprese di grande dimensione.

Il restante cluster è formato dalle istanze con valori più bassi del Valore della Produzione e risulta essere fittamente concentrato. Analogamente a come proceduto precedentemente, è stata applicata un'ulteriore clusterizzazione all'interno del cluster 0, per poter meglio separare le istanze.

Al fine di individuare il valore ottimale del parametro K , è stato nuovamente utilizzato il *Loop Parameters*. I primi risultati ordinati in ordine crescente sono stati riportati in Figura 3.4:

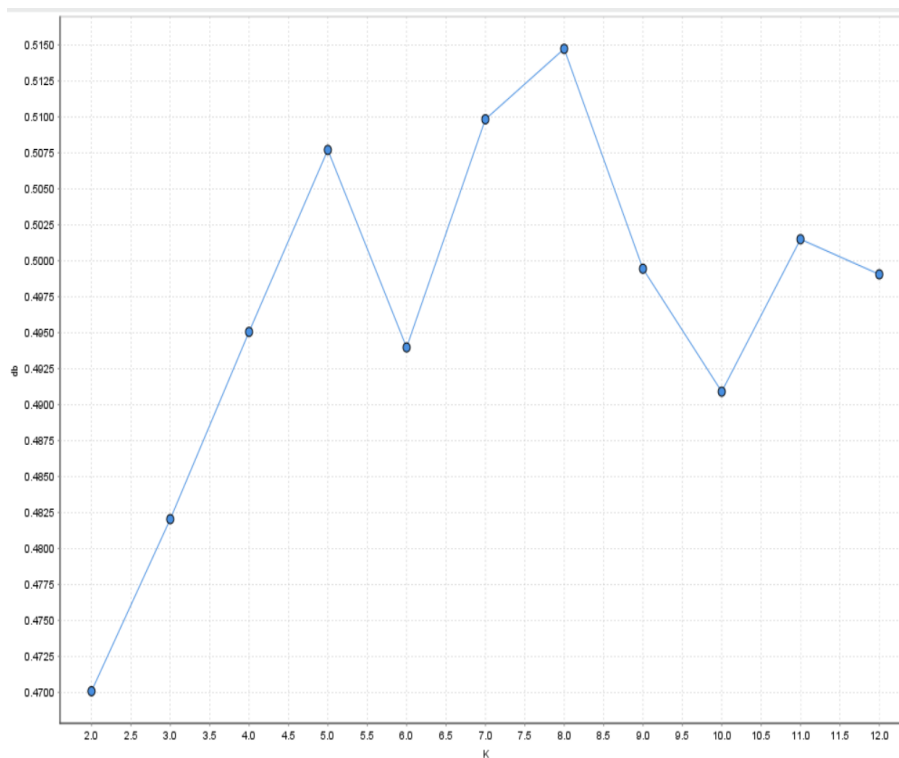


Figura 3.4: Clusterizzazione con K-MEANS

Analogamente al primo, anche in questo caso il valore ottimale è 2.

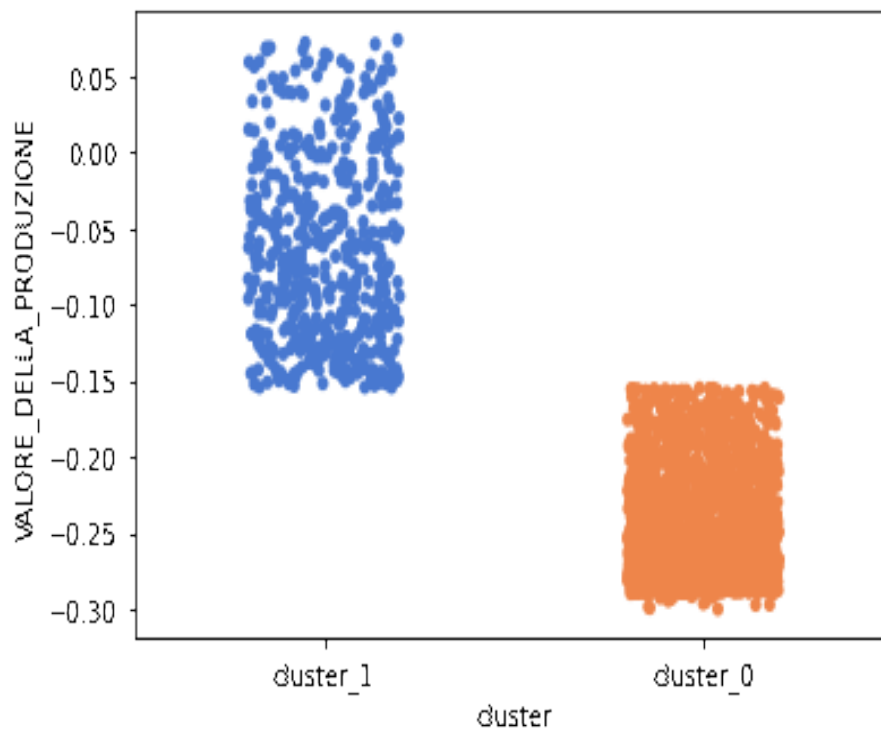


Figura 3.5: Clusterizzazione con K-MEANS.2

La tabella 3.4 descrive i singoli cluster:

Cluster	Numero Imprese	Centroide
cluster_0	1836	7316,66
cluster_1	506	24677,83

Tabella 3.4: Caratteristiche dei cluster

L'ulteriore clusterizzazione ha permesso la formazione di due gruppi che mostrano caratteristiche differenti. Il cluster 0, con valori della produzione minori,

è formato dalle piccole imprese. Il cluster 1, invece, rappresenta le imprese di medie dimensioni.

Nel complesso, il risultato della *Cluster Analysis* effettuata con entrambi le tipologie di algoritmi è mostrato nella tabella 3.5:

Cluster	Nome Cluster	Numero Imprese	Centroide
Noise	Imprese Outliers	119	
cluster_0	Piccole imprese	1836	7316,66
cluster_1	Medie imprese	506	24677,83
cluster_2	Grandi imprese	344	65187,22

Tabella 3.5: Caratteristiche dei cluster

Capitolo 3

4.1 Composizione del dataset

A seguito della divisione delle istanze in gruppi omogenei, il dataset è stato costruito nel seguente modo: per ogni impresa sono state inserite come variabili indipendenti le variabili di bilancio e le variabili macroeconomiche a partire dall'anno t_i e come variabile dipendente l'Ebitda Margin dell'anno t_{i+1} .

TRAINING SET

VARIABILI INDIPENDENTI	VARIABILE DIPENDENTE
variabili 2014	Ebitda Margin 2015
variabili 2015	Ebitda Margin 2016
variabili 2016	Ebitda Margin 2017
variabili 2017	Ebitda Margin 2018

Tabella 4.1: Training set

In questo modo si creano quattro ID non univoci per la stessa impresa, con valori differenti delle singole variabili.

TEST SET

VARIABILI INDIPENDENTI	VARIABILE DIPENDENTE
variabili 2018	Ebitda Margin 2019

Tabella 4.2: Test set

A differenza del training set, il test set si compone di ID univoci, prendendo a riferimento solo il 2018 come anno delle variabili indipendenti.

La costruzione del dataset attraverso la modalità appena esposta ha come fine quello di prevedere i valori dell'Ebitda Margin di un determinato anno, attraverso i valori delle variabili dell'anno precedente a quello di interesse.

I risultati derivanti dalla precedente clusterizzazione sono stati riformulati alla luce della nuova struttura del dataset, ottenendo la seguente composizione:

SPLIT	ISTANZE	% TOTALE	PICCOLE IMPRESE	MEDIE IMPRESE	GRANDI IMPRESE	OUTLIERS
Training set	11220	80%	7344	2024	1376	476
Test set	2805	20%	1836	506	344	119

Tabella 4.3: Composizione del Training e Test set

4.2 Random Forest per Features Selection

Per *Features Selection* si intende il metodo di selezione di una parte delle variabili indipendenti al fine di ridurre la dimensionalità del problema e permettere di costruire modelli di *Machine Learning* più semplici e accurati.

L'algoritmo utilizzato per la *Features Selection* è stato quello del Random Forest.

4.2.1 Funzionamento

Il *Random Forest* è un algoritmo utilizzabile per regressione, che crea un insieme di alberi decisionali indipendenti. Al fine di garantire la de-correlazione degli alberi:

- si addestra ogni albero su un campione *bootstrap* dell'intero Training Set.
- per ogni albero, ad ogni split, si selezionano randomicamente un set di predittori dall'intero insieme dei variabili.

I due principali parametri da ottimizzare sono:

- NUMERO DI ALBERI: numero di alberi totali da costruire.
- PROFONDITÀ MASSIMA: profondità massima di ogni albero dell'insieme creato.

L'Algoritmo del Random Forest è stato utilizzato in *10-Fold Cross Validation* solo sul Training Test.

10-Fold Cross Validation

La *Cross Validation* è una tecnica utilizzata nel Machine Learning per valutare la performance di un modello. La *10-Fold Cross Validation* si basa sul presupposto che il dataset di riferimento, ad ogni iterazione, venga diviso in dieci sottoinsiemi casuali. Uno dei sottoinsiemi è considerabile come test, mentre i restanti nove come training. Ad ogni iterazione si ottiene un errore di test. La stima dell'errore di regressione del modello deriva dalla media dei dieci errori ottenuti dalle singole iterazioni. Un modello che presenta congiuntamente un errore e una varianza dell'errore bassi, è considerabile il migliore, perché stabile e capace di generalizzare in presenza di nuove istanze.

4.2.2 Metriche

La metrica utilizzata per valutare la bontà della predizione è stata quella del *Root Mean Squared Error* (RMSE). Esso calcola la radice quadrata della media dei residui al quadrato, dove per residui si intende la differenza tra il valore predetto e il valore reale che la variabile dipendente assume per ogni istanza.

Di seguito la formula dell'RMSE:

$$RMSE = \sqrt{\sum_{i=1}^N \frac{(\hat{y}_i - y_i)^2}{N}} \quad (4.1)$$

4.2.3 Analisi

All'interno di ogni cluster, è stata introdotta un'ulteriore suddivisione per Forma Giuridica, ovvero S.p.a, S.r.l o Società Cooperativa.

Per ogni sottoinsieme è stato quindi implementato il *Random Forest* in CV, estraendo i pesi di ciascuna feature, per poter selezionare le più importanti.

I risultanti sono riportati in Tabella 4.4-4.6.

Tabella 4.4: PICCOLE IMPRESE

SPA	weights	SRL	weights	SOCIETÀ COOPERATIVE	weights
TOT_DISP_LIQUIDE	0.3024	TOTALE.IMMOBILIZZAZIONI	0.4892	TOTALE.IMMOBILIZZAZIONI	0.7323
CREDITLA.BREVE	0.2999	TOT_DISP_LIQUIDE	0.4112	DIPENDENTI	0.2391
DIPENDENTI	0.2188	VALORE.DELLA.PRODUZIONE	0.2118	DEBITLOLTRE	0.1723
TOTALE.IMMOBILIZZAZIONI	0.2034	INFLAZIONE	0.0629	TOTALE.RIMANENZE	0.0733
COSTL.PRODUZIONE	0.1036	INDEX.PREZZI.CONSUMO	0.0397	TOT_DISP_LIQUIDE	0.0211
VALORE.DELLA.PRODUZIONE	0.0712	PIL.PRO.CAPITE	0.0293	CREDITLOLTRE	0.0208
PIL.PRO.CAPITE	0.0328	CREDITLOLTRE	0.0192	INDEX.PREZZI.CONSUMO	0.0206
CREDITLOLTRE	0.0317	DEBITLOLTRE	0.0188	CONSUMER.CONF.INDEX	0.0138
ISIC	0.0313	ISIC	0.0125	INFLAZIONE	0.0085
INFLAZIONE	0.0259	BUSINESS.CONF.INDEX	0.0063	PIL.PRO.CAPITE	0.0085
TOTALE.RIMANENZE	0.0175	CREDITLA.BREVE	0.0011	DEBITLA.BREVE	0.0017
INDEX.PREZZI.CONSUMO	0.0103	GRADO.APERTURA.COMMERCIALE	0.0000	CREDITLA.BREVE	0.0017
CONSUMER.CONF.INDEX	0.0066	TOTALE.RIMANENZE	0.0000	COSTL.PRODUZIONE	0.0013
BUSINESS.CONF.INDEX	0.0041	DIPENDENTI	0.0000	ISIC	0.0012
GRADO.APERTURA.COMMERCIALE	0.0029	CONSUMER.CONF.INDEX	0.0000	BUSINESS.CONF.INDEX	0.0000
DEBITLOLTRE	0.0017	DEBITLA.BREVE	0.0000	GRADO.APERTURA.COMMERCIALE	0.0000
DEBITLA.BREVE	0.0000	COSTL.PRODUZIONE	0.0000	VALORE.DELLA.PRODUZIONE	0.0000

Tabella 4.5: MEDIE IMPRESE

SPA	weights	SRL	weights	SOCIETÀ COOPERATIVE	weights
TOTALE.IMMOBILIZZAZIONI	0.1693	TOT_DISP_LIQUIDE	0.4756	TOTALE.IMMOBILIZZAZIONI	0.5158
TOT_DISP_LIQUIDE	0.1611	TOTALE.IMMOBILIZZAZIONI	0.3440	DIPENDENTI	0.3242
INDEX_PREZZI_CONSUMO	0.0457	DIPENDENTI	0.2404	CREDITLA_BREVE	0.1735
INFLAZIONE	0.0298	CREDITLA_BREVE	0.0959	TOT_DISP_LIQUIDE	0.1674
CREDITI_OLTRE	0.0282	INFLAZIONE	0.0692	TOTALE_RIMANENZE	0.0814
VALORE_DELLA_PRODUZIONE	0.0257	TOTALE_RIMANENZE	0.0621	DEBITI_OLTRE	0.0679
DEBITI_OLTRE	0.0236	CREDITI_OLTRE	0.0589	VALORE_DELLA_PRODUZIONE	0.0604
GRADO_APERTURA_COMMERCIALE	0.0134	ISIC	0.0268	COSTI_PRODUZIONE	0.0329
CREDITLA_BREVE	0.0098	VALORE_DELLA_PRODUZIONE	0.0196	INDEX_PREZZI_CONSUMO	0.0310
DIPENDENTI	0.0095	INDEX_PREZZI_CONSUMO	0.0140	INFLAZIONE	0.0232
BUSINESS_CONF_INDEX	0.0093	PIL_PRO_CAPITE	0.0039	CREDITI_OLTRE	0.0187
PIL_PRO_CAPITE	0.0060	GRADO_APERTURA_COMMERCIALE	0.0037	ISIC	0.0103
ISIC	0.0056	CONSUMER_CONF_INDEX	0.0035	CONSUMER_CONF_INDEX	0.0050
TOTALE_RIMANENZE	0.0032	DEBITI_OLTRE	0.0000	PIL_PRO_CAPITE	0.0030
COSTI_PRODUZIONE	0.0011	BUSINESS_CONF_INDEX	0.0000	DEBITLA_BREVE	0.0024
DEBITLA_BREVE	0.0000	COSTI_PRODUZIONE	0.0000	GRADO_APERTURA_COMMERCIALE	0.0020
CONSUMER_CONF_INDEX	0.0000	DEBITLA_BREVE	0.0000	BUSINESS_CONF_INDEX	0.0019

Tabella 4.6: GRANDI IMPRESE

SPA	weights	SRL	weights	SOCIETÀ COOPERATIVE	weights
TOT_DISP_LIQUIDE	0.3323	TOTALE.IMMOBILIZZAZIONI	0.1989	TOTALE.IMMOBILIZZAZIONI	0.2022
TOTALE.IMMOBILIZZAZIONI	0.3071	TOT_DISP_LIQUIDE	0.1540	TOT_DISP_LIQUIDE	0.1543
DIPENDENTI	0.2334	INDEX_PREZZI_CONSUMO	0.0241	INDEX_PREZZI_CONSUMO	0.0240
VALORE_DELLA_PRODUZIONE	0.1035	BUSINESS_CONF_INDEX	0.0216	BUSINESS_CONF_INDEX	0.0217
INFLAZIONE	0.0722	ISIC	0.0138	ISIC	0.0141
CREDITLA_BREVE	0.0700	INFLAZIONE	0.0123	INFLAZIONE	0.0125
CREDITI_OLTRE	0.0602	GRADO_APERTURA_COMMERCIALE	0.0123	GRADO_APERTURA_COMMERCIALE	0.0123
ISIC	0.0348	PIL_PRO_CAPITE	0.0105	PIL_PRO_CAPITE	0.0107
DEBITI_OLTRE	0.0212	VALORE_DELLA_PRODUZIONE	0.0060	VALORE_DELLA_PRODUZIONE	0.0063
PIL_PRO_CAPITE	0.0104	CREDITLA_BREVE	0.0045	CREDITLA_BREVE	0.0047
INDEX_PREZZI_CONSUMO	0.0063	CREDITI_OLTRE	0.0024	CREDITI_OLTRE	0.0025
GRADO_APERTURA_COMMERCIALE	0.0055	DEBITI_OLTRE	0.0018	DEBITI_OLTRE	0.0017
BUSINESS_CONF_INDEX	0.0054	TOTALE_RIMANENZE	0.0000	TOTALE_RIMANENZE	0.0000
CONSUMER_CONF_INDEX	0.0038	DEBITLA_BREVE	0.0000	DEBITLA_BREVE	0.0000
TOTALE_RIMANENZE	0.0037	CONSUMER_CONF_INDEX	0.0000	CONSUMER_CONF_INDEX	0.0000
COSTI_PRODUZIONE	0.0000	COSTI_PRODUZIONE	0.0000	COSTI_PRODUZIONE	0.0000
DEBITLA_BREVE	0.0000	DIPENDENTI	0.0000	DIPENDENTI	0.0000

Per ogni casistica le variabili sono disposte in ordine decrescente per il peso associato. Sono state quindi selezionate le prime dieci di ognuna sulle diciotto totali.

4.3 Loop Attributes Subset

Dopo la scelta delle prime dieci variabili per ogni cluster, si è convenuto procedere con un'ulteriore selezione delle *features*. Nello specifico, attraverso l'operatore *Loop Attributes Subset*, sono state create, all'interno di ogni casistica, tutte le possibili combinazioni di variabili con i relativi valori dell'RMSE e della *Standard Deviation* dell'RMSE.

Per poter estrarre le combinazioni, sono state utilizzati tre differenti algoritmi, inseriti all'interno della *Cross Validation*, quali: Decision Trees, Random Forest, Support Vector Regression. Con ognuno, per ogni combinazione dei variabili creatasi, oltre ai corrispondenti valori dell'RMSE e della *Standard Deviation* dell'RMSE, è presente la combinazione di parametri dell'algoritmo utilizzato. Tale procedura è finalizzata ad individuare la combinazione di features e parametri a cui corrisponde il valore dell'RMSE minore.

Per avere un'analisi il più accurata possibile, è stata effettuata un'ulteriore diversificazione all'interno dei cluster già formati. In particolare, per ogni casistica è stata introdotta la divisione territoriale in Nord, Centro, Sud/Isole. Un'analisi approfondita della redditività delle imprese non può prescindere, infatti, dall'ambito territoriale e settoriale in cui l'impresa stessa opera. Da qui la necessità di costruire modelli di previsioni che considerino la singola azienda in relazione all'ambito dove la stessa è collocata.

La Tabella 4.7 mostra i nuovi gruppi formatisi:

CLUSTER	FORMA GIURIDICA	AREA
PICCOLE IMPRESE	SPA	Nord Centro Sud/Isole
	SRL	Nord Centro Sud/Isole
	Società Cooperative	Nord Centro Sud/Isole
MEDIE IMPRESE	SPA	Nord Centro Sud/Isole
	SRL	Nord Centro Sud/Isole
	Società Cooperative	Nord Centro Sud/Isole
GRANDI IMPRESE	SPA	Nord Centro Sud/Isole
	SRL	Nord Centro Sud/Isole
	Società Cooperative	Nord Centro Sud/Isole

Tabella 4.7: Nuovi cluster

In definitiva, le imprese presenti nel dataset originali sono state suddivise in base a una logica dimensionale, territoriale e di forma giuridica.

4.4 Decision Tree

L'albero decisionale è un algoritmo di *machine learning* utilizzabile in classificazione e regressione.

4.4.1 Funzionamento

Un *decision tree* ha una struttura ad albero formata da:

- Nodo radice: nodo senza collegamenti in entrata e zero o più in uscita.
- Nodi interni: nodi che hanno un collegamento in entrata e due o più in uscita.
- Foglie o nodi terminali: nodi che hanno un collegamento in entrata e nessuno in uscita.

Ad ogni step, uno degli attributi viene selezionato e le istanze partizionate in differenti nodi figli. Il criterio che determina quale attributo venga scelto è il criterio di split. Nel caso di regressione il criterio utilizzato è il *Sum of Squared Errors*, ovvero la differenza tra i valori osservati e i valori predetti. L'attributo selezionato sarà quello che minimizza tale valore.

In questo caso, l'unico parametro da ottimizzare è la profondità massima dell'albero. All'aumentare della profondità, aumenta il numero di split e la capacità, quindi, di estrapolare informazione dai dati.

4.4.2 Analisi

Il parametro 'profondità massima' è stato l'unico ottimizzato nel loop di ogni singola casistica, secondo il seguente range:

Parametro	Min	Max	Steps	Scale
Maximal depth	10	200	10	logarithmic

Tabella 4.8: Setting dei parametri del Decision Tree

Di seguito sono riportati i primi cinque risultati di ogni loop, disposti in ordine crescente in relazione all'RMSE. Vengono presi in considerazione, quindi, le combinazioni di variabili e parametri con l'errore più basso possibile.

PICCOLE IMPRESE

SPA NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.825572	0.14492	10.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI	0.825582	0.11051	10.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, ISIC	0.825686	0.11437	10.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, PIL.PRO.CAPITE, ISIC	0.825986	0.11456	10.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, INFLAZIONE, ISIC	0.827221	0.11937	10.0

Tabella 4.9: Tuning dei parametri del Decision Tree: piccole imprese Spa Nord

SPA CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
COSTI.PRODUZIONE, CREDITI.A.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI	1,6128	2,5955	25.0
COSTI.PRODUZIONE, CREDITI.A.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI	1,6128	2,5955	110.0
COSTI.PRODUZIONE, CREDITI.A.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI	1,6128	2,5955	45.0
COSTI.PRODUZIONE, CREDITI.A.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI	1,6128	2,5955	18.0
COSTI.PRODUZIONE, CREDITI.A.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI	1,6128	2,5955	148.0

Tabella 4.10: Tuning dei parametri del Decision Tree: piccole imprese Spa Centro

SPA SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, CREDITI.OLTRE, DIPENDENTI	0.54129	0.14661	10.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, CREDITI.OLTRE, DIPENDENTI	0.54129	0.14856	13.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, CREDITI.OLTRE, DIPENDENTI	0.54224	0.14888	110.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, CREDITI.OLTRE, DIPENDENTI	0.54224	0.14888	45.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, CREDITI.OLTRE, DIPENDENTI	0.54224	0.14888	18.0

Tabella 4.11: Tuning dei parametri del Decision Tree: piccole imprese Spa Sud/Isole

SRL NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
TOT_DISP_LIQUIDE	0.97703	0.20307	10.0
TOT_DISP_LIQUIDE	0.98358	0.19866	13.0
TOT_DISP_LIQUIDE, BUSINESS_CONF_INDEX	0.98436	0.20488	10.0
TOT_DISP_LIQUIDE	0.98498	0.19876	60.0
TOT_DISP_LIQUIDE	0.98498	0.19876	45.0

Tabella 4.12: Tuning dei parametri del Decision Tree: piccole imprese Srl Nord

SRL CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITLOLTRE, INFLAZIONE	0.93121	0.24141	10.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITLOLTRE, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.93836	0.23749	10.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITLOLTRE, PIL_PRO_CAPITE, INDEX.PREZZI.CONSUMO	0.94849	0.28327	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITLOLTRE, BUSINESS_CONF_INDEX	0.94927	0.24452	10.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITLOLTRE, PIL_PRO_CAPITE	0.94976	0.27377	10.0

Tabella 4.13: Tuning dei parametri del Decision Tree: piccole imprese Srl

Centro

SRL SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP.LIQUIDE, BUSINESS.CONF.INDEX, INFLAZIONE	0.74145	0.14654	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP.LIQUIDE, INFLAZIONE	0.74284	0.14646	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP.LIQUIDE, DEBITI.OLTRE, INFLAZIONE	0.74320	0.14440	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE TOT_DISP.LIQUIDE, INFLAZIONE	0.74399	0.14817	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP.LIQUIDE, DEBITI.OLTRE, BUSINESS.CONF.INDEX, INFLAZIONE	0.74472	0.14448	10.0

Tabella 4.14: Tuning dei parametri del Decision Tree: piccole imprese Srl Sud/Centro

SOCIETÀ COOPERATIVE NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, CONSUMER.CONF.INDEX, INFLAZIONE	0.24886	0.09676	13.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, CONSUMER.CONF.INDEX, INFLAZIONE	0.24980	0.09496	10.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, CONSUMER.CONF.INDEX, INFLAZIONE	0.25000	0.09596	18.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, CONSUMER.CONF.INDEX, INFLAZIONE	0.25000	0.09596	33.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, CONSUMER.CONF.INDEX, INFLAZIONE	0.25000	0.09596	110.0

Tabella 4.15: Tuning dei parametri del Decision Tree: piccole imprese Società Cooperative Nord

SOCIETÀ COOPERATIVE CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
CREDITI_OLTRE, DIPENDENTI	0.11824	0.06202	18.0
CREDITI_OLTRE, DIPENDENTI	0.11824	0.06202	45.0
CREDITI_OLTRE, DIPENDENTI	0.11824	0.06202	33.0
CREDITI_OLTRE, DIPENDENTI	0.11824	0.06202	110.0
CREDITI_OLTRE, DIPENDENTI	0.11824	0.06202	25.0

Tabella 4.16: Tuning dei parametri del Decision Tree: piccole imprese Società Cooperative Centro

SOCIETÀ COOPERATIVE SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DIPENDENTI, PIL.PRO.CAPITE, INDEX.PREZZI.CONSUMO, CONSUMER.CONF.INDEX	0.36063	0.21355	13.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DIPENDENTI, PIL.PRO.CAPITE, INDEX.PREZZI.CONSUMO, CONSUMER.CONF.INDEX	0.36100	0.21327	110.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DIPENDENTI, PIL.PRO.CAPITE, INDEX.PREZZI.CONSUMO, CONSUMER.CONF.INDEX	0.36100	0.21327	148.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DIPENDENTI, PIL.PRO.CAPITE, INDEX.PREZZI.CONSUMO, CONSUMER.CONF.INDEX	0.36100	0.21327	45.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DIPENDENTI, PIL.PRO.CAPITE, INDEX.PREZZI.CONSUMO, CONSUMER.CONF.INDEX	0.36100	0.21327	18.0

Tabella 4.17: Tuning dei parametri del Decision Tree: piccole imprese Società Cooperative Sud/Isole

MEDIE IMPRESE

SPA NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DEBITI_OLTRE, DIPENDENTI, GRADO_APERTURA_COMMERCIALE	0.92847	0.31439	10.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DEBITI_OLTRE, DIPENDENTI, INDEX_PREZZI_CONSUMO	0.92859	0.31429	10.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DEBITI_OLTRE, DIPENDENTI, INDEX_PREZZI_CONSUMO, INFLAZIONE	0.92897	0.31354	10.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DEBITI_OLTRE, DIPENDENTI, GRADO_APERTURA_COMMERCIALE, INFLAZIONE	0.92904	0.31364	10.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DEBITI_OLTRE, DIPENDENTI, INDEX_PREZZI_CONSUMO, GRADO_APERTURA_COMMERCIALE	0.92928	0.31339	10.0

Tabella 4.18: Tuning dei parametri del Decision Tree: medie imprese Spa Nord

SPA CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DIPENDENTI	0.23898	0.11060	110.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DIPENDENTI	0.23898	0.11060	81.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DIPENDENTI	0.23898	0.11060	10.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DIPENDENTI	0.23898	0.11060	45.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DIPENDENTI	0.23898	0.11060	60.0

Tabella 4.19: Tuning dei parametri del Decision Tree: medie imprese Spa

Centro

SPA SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
DEBITI_OLTRE	1,2792	0.92216	110.0
DEBITI_OLTRE	1,2792	0.92216	60.0
DEBITI_OLTRE	1,2792	0.92216	148.0
DEBITI_OLTRE	1,2792	0.92216	45.0
DEBITI_OLTRE	1,2792	0.92216	18.0

Tabella 4.20: Tuning dei parametri del Decision Tree: medie imprese Spa Sud/Isole

SRL NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE DELLA PRODUZIONE, CREDITI_OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI	0.84344	0.18399	10.0
VALORE DELLA PRODUZIONE, CREDITI_OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, INDEX_PREZZI_CONSUMO	0.84464	0.17824	10.0
VALORE DELLA PRODUZIONE, CREDITI_OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, INDEX_PREZZI_CONSUMO, INFLAZIONE	0.84570	0.18047	10.0
VALORE DELLA PRODUZIONE, CREDITI_OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, INFLAZIONE	0.84688	0.17968	10.0
VALORE DELLA PRODUZIONE, CREDITI_OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, INDEX_PREZZI_CONSUMO, ISIC	0.85286	0.18343	10.0

Tabella 4.21: Tuning dei parametri del Decision Tree: medie imprese Srl Nord

SRL CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITLA.BREVE, TOT_DISP_LIQUIDE, DIPENDENTI	0.53632	0.20540	13.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITLA.BREVE, TOT_DISP_LIQUIDE, DIPENDENTI	0.53718	0.23184	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITLA.BREVE, TOT_DISP_LIQUIDE, DIPENDENTI	0.53927	0.20347	25.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITLA.BREVE, TOT_DISP_LIQUIDE, DIPENDENTI	0.53927	0.20347	45.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITLA.BREVE, TOT_DISP_LIQUIDE, DIPENDENTI	0.53927	0.20347	110.0

Tabella 4.22: Tuning dei parametri del Decision Tree: medie imprese Srl Centro

SRL SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE.DELLA.PRODUZIONE, CREDITI.OLTRE, DIPENDENTI	0.59321	0.11869	10.0
VALORE.DELLA.PRODUZIONE, CREDITI.OLTRE, DIPENDENTI	0.59582	0.13066	13.0
VALORE.DELLA.PRODUZIONE, CREDITI.OLTRE, DIPENDENTI	0.59950	0.13891	25.0
VALORE.DELLA.PRODUZIONE, CREDITI.OLTRE, DIPENDENTI	0.59950	0.13891	110.0
VALORE.DELLA.PRODUZIONE, CREDITI.OLTRE, DIPENDENTI	0.59950	0.13891	18.0

Tabella 4.23: Tuning dei parametri del Decision Tree: medie imprese Srl Sud/Isole

SOCIETÀ COOPERATIVE NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOTALE RIMANENZE, DIPENDENTI	0.15671	0.07269	16.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOTALE RIMANENZE, DIPENDENTI	0.15671	0.07269	45.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOTALE RIMANENZE, DIPENDENTI	0.15671	0.07269	95.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOTALE RIMANENZE, DIPENDENTI	0.15671	0.07269	13.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOTALE RIMANENZE, DIPENDENTI	0.15671	0.07269	12.0

Tabella 4.24: Tuning dei parametri del Decision Tree: medie imprese Società Cooperative Nord

SOCIETÀ COOPERATIVE CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
DEBITI OLTRE, DIPENDENTI, INFLAZIONE	0.10691	0.07329	21.0
DEBITI OLTRE, DIPENDENTI, INFLAZIONE	0.10691	0.07329	29.0
DEBITI OLTRE, DIPENDENTI, INFLAZIONE	0.10691	0.07329	12.0
DEBITI OLTRE, DIPENDENTI, INFLAZIONE	0.10691	0.07329	10.0
DEBITI OLTRE, DIPENDENTI, INFLAZIONE	0.10691	0.07329	25.0

Tabella 4.25: Tuning dei parametri del Decision Tree: medie imprese Società Cooperative Centro

SOCIETÀ COOPERATIVE SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, CREDITI A BREVE, TOT. DISP. LIQUIDE, INDEX PREZZI CONSUMO, INFLAZIONE	0.24865	0.14625	21.0
VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, CREDITI A BREVE, TOT. DISP. LIQUIDE, INDEX PREZZI CONSUMO, INFLAZIONE	0.24865	0.14625	45.0
VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, CREDITI A BREVE, TOT. DISP. LIQUIDE, INDEX PREZZI CONSUMO, INFLAZIONE	0.24865	0.14625	95.0
VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, CREDITI A BREVE, TOT. DISP. LIQUIDE, INDEX PREZZI CONSUMO, INFLAZIONE	0.24865	0.14625	10.0
VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, CREDITI A BREVE, TOT. DISP. LIQUIDE, INDEX PREZZI CONSUMO, INFLAZIONE	0.24865	0.14625	13.0

Tabella 4.26: Tuning dei parametri del Decision Tree: medie imprese Società Cooperative Sud/Isole

GRANDI IMPRESE

SPA NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, TOT. DISP. LIQUIDE, DEBITI OLTRE, DIPENDENTI	0.81923	0.16256	10.0
TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, TOT. DISP. LIQUIDE, DEBITI OLTRE, DIPENDENTI, ISIC	0.82108	0.15385	10.0
TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, TOT. DISP. LIQUIDE, DEBITI OLTRE, DIPENDENTI, PIL PRO. CAPITE	0.82133	0.15360	10.0
TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, TOT. DISP. LIQUIDE, DEBITI OLTRE, DIPENDENTI, PIL PRO. CAPITE, ISIC	0.82142	0.15295	10.0
TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, TOT. DISP. LIQUIDE, DEBITI OLTRE, DIPENDENTI, PIL PRO. CAPITE, INFLAZIONE, ISIC	0.82299	0.15952	10.0

Tabella 4.27: Tuning dei parametri del Decision Tree: grandi imprese Spa Nord

SPA CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, PIL.PRO.CAPITE	0.39455	0.28618	18.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, PIL.PRO.CAPITE	0.39455	0.28618	13.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, PIL.PRO.CAPITE	0.39455	0.28618	148.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, PIL.PRO.CAPITE	0.39455	0.28618	45.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, PIL.PRO.CAPITE	0.39455	0.28618	25.0

Tabella 4.28: Tuning dei parametri del Decision Tree: grandi imprese Spa Centro

SPA SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.LA.BREVE, DIPENDENTI	0.68535	0.20427	13.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.LA.BREVE, DIPENDENTI	0.68535	0.20427	110.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.LA.BREVE, DIPENDENTI	0.68535	0.20427	45.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.LA.BREVE, DIPENDENTI	0.68535	0.20427	18.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.LA.BREVE, DIPENDENTI	0.68535	0.20427	25.0

Tabella 4.29: Tuning dei parametri del Decision Tree: grandi imprese Spa Sud/Isole

SRL NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE, BUSINESS_CONF_INDEX	1027227	0.25184	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE, ISIC	1027484	0.25459	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE, PIL_PRO_CAPITE, ISIC	1027971	0.25521	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE	1028296	0.24449	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE, PIL_PRO_CAPITE	1029831	0.25255	10.0

Tabella 4.30: Tuning dei parametri del Decision Tree: grandi imprese Srl Nord

SRL CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, GRADO.APERTURA.COMMERCIALE	0.24377	0.11498	18.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, GRADO.APERTURA.COMMERCIALE	0.24377	0.11498	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, GRADO.APERTURA.COMMERCIALE	0.24377	0.11498	45.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, GRADO.APERTURA.COMMERCIALE	0.24377	0.11498	13.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, GRADO.APERTURA.COMMERCIALE	0.24377	0.11498	25.0

Tabella 4.31: Tuning dei parametri del Decision Tree: grandi imprese Srl Centro

SRL SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, GRADO.APERTURA.COMMERCIALE	0.69871	0.18889	33.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, GRADO.APERTURA.COMMERCIALE	0.69871	0.18889	45.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, GRADO.APERTURA.COMMERCIALE	0.69871	0.18889	25.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, GRADO.APERTURA.COMMERCIALE	0.69871	0.18889	110.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, GRADO.APERTURA.COMMERCIALE	0.69871	0.18889	18.0

Tabella 4.32: Tuning dei parametri del Decision Tree: grandi imprese Srl Sud/Isole

SOCIETÀ COOPERATIVE NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE	0.16827	0.06917	148.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE	0.16827	0.06917	110.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE	0.16827	0.06917	25.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE	0.16827	0.06917	13.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE	0.16827	0.06917	18.0

Tabella 4.33: Tuning dei parametri del Decision Tree: grandi imprese Società Cooperative Nord

Le restanti due casistiche, Società Cooperative Centro e Società Cooperative Sud/Isole, non contengono un numero di istanze tale da poter utilizzare l'operatore del *Loop Attribute Subset*.

4.5 Random Forest

4.5.1 Analisi

Analogamente a come proceduto con il Decision Tree, l'utilizzo del Random Forest è stato propedeutico ad individuare combinazioni di attributi che meglio riescano a spiegare il fenomeno analizzato. In questo caso i parametri presi in considerazione sono:

- Profondità Massima: profondità massima di ogni albero costruito.
- Numero di alberi: numero totale di alberi generati.

Il setting dei due parametri è esposto in tabella 4.34:

Parametri	Min	Max	Steps	Scale
Maximal depth	10	100	5	logarithmic
Number of trees	100	1000	3	logarithmic

Tabella 4.34: Setting dei parametri del Random Forest

Di seguito sono riportati i primi cinque risultati di ogni loop, disposti in ordine crescente in relazione all'RMSE.

PICCOLE IMPRESE

SPA NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP.LIQUIDE, DIPENDENTI	0.66595	0.13560	100.0	464.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, CREDITLOLTRE, TOT_DISP.LIQUIDE, DIPENDENTI	0.66649	0.12585	100.0	215.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP.LIQUIDE, DIPENDENTI	0.66658	0.13387	100.0	1000.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP.LIQUIDE, DIPENDENTI	0.66704	0.13352	25.0	1000.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP.LIQUIDE, DIPENDENTI	0.66752	0.13503	63.0	1000.0

Tabella 4.35: Tuning dei parametri del Random Forest: piccole imprese Spa Nord

SPA CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, CREDITLA.BREVE, DIPENDENTI	1,465,374	2,463,430	25.0	100.0
VALORE.DELLA.PRODUZIONE, CREDITLA.BREVE, CREDITLOLTRE, DIPENDENTI	1,4662	2,4648	40.0	100.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, CREDITLA.BREVE, CREDITLOLTRE, DIPENDENTI	1,4665	2,4481	40.0	215.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, CREDITLA.BREVE, CREDITLOLTRE, DIPENDENTI	1,4712	2,4471	16.0	100.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, CREDITLA.BREVE, CREDITLOLTRE, TOT_DISP.LIQUIDE, DIPENDENTI	1,4715	2,5640	16.0	100.0

Tabella 4.36: Tuning dei parametri del Random Forest: piccole imprese Spa Centro

SPA SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
COSTLPRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, DIPENDENTI	0.43791	0.12883	16.0	464.0
COSTLPRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, DIPENDENTI	0.43940	0.12859	25.0	464.0
COSTLPRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, DIPENDENTI	0.43952	0.12780	16.0	1000.0
COSTLPRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, DIPENDENTI	0.43975	0.12967	40.0	100.0
COSTLPRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, DIPENDENTI	0.44045	0.12490	100.0	100.0

Tabella 4.37: Tuning dei parametri del Random Forest: piccole imprese Spa Sud/Isole

SRL NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.90392	0.23683	100.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE	0.90779	0.23797	40.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.90863	0.24128	63.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.90879	0.23886	100.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.90921	0.24452	40.0	215.0

Tabella 4.38: Tuning dei parametri del Random Forest: piccole imprese Srl Nord

SRL CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.80346	0.27986	16.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.80856	0.27176	63.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.81095	0.27995	25.0	464.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.81125	0.27965	100.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.81128	0.27330	25.0	464.0

Tabella 4.39: Tuning dei parametri del Random Forest: piccole imprese Srl
Centro

SRL SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.62885	0.12913	100.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.63107	0.12898	16.0	1000.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.63234	0.12978	63.0	1000.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.63248	0.12700	63.0	464.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.63254	0.13220	25.0	100.0

Tabella 4.40: Tuning dei parametri del Random Forest: piccole imprese Srl
Sud/Isole

SOCIETÀ COOPERATIVE NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI	0.23189	0.09863	100.0	215.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DEBITI.OLTRE, DIPENDENTI	0.23266	0.10241	40.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITI.OLTRE, DIPENDENTI	0.23300	0.10640	100.0	215.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DEBITI.OLTRE, DIPENDENTI	0.23396	0.10276	100.0	1000.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DEBITI.OLTRE, DIPENDENTI	0.23403	0.10416	40.0	464.0

Tabella 4.41: Tuning dei parametri del Random Forest: piccole imprese Società Cooperative Nord

SOCIETÀ COOPERATIVE CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DEBITI.OLTRE, DIPENDENTI, INDEX_PREZZI_CONSUMO	0.14226	0.13763	25.0	100.0
TOT_DISP_LIQUIDE, DEBITI.OLTRE, DIPENDENTI, INFLAZIONE	0.14228	0.12224	40.0	100.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI	0.14229	0.12949	16.0	100.0
TOTALE.IMMOBILIZZAZIONI, DEBITI.OLTRE, DIPENDENTI	0.14272	0.11732	16.0	100.0
TOTALE.IMMOBILIZZAZIONI, DEBITI.OLTRE, DIPENDENTI, CONSUMER.CONF.INDEX	0.14282	0.12670	100.0	100.0

Tabella 4.42: Tuning dei parametri del Random Forest: piccole imprese Società Cooperative Centro

SOCIETÀ COOPERATIVE SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP_LIQUIDE, DEBITI.OLTR	0.31286	0.11735	63.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.31355	0.11453	10.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.31421	0.11436	25.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.31461	0.11097	40.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.31476	0.11440	16.0	215.0

Tabella 4.43: Tuning dei parametri del Random Forest: piccole imprese Società Cooperative Sud/Isole

MEDIE IMPRESE

SPA NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.87940	0.23054	100.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.88283	0.22974	63.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.88369	0.23387	25.0	215.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.88382	0.24301	16.0	1000.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.88625	0.24028	25.0	1000.0

Tabella 4.44: Tuning dei parametri del Random Forest: medie imprese Spa Nord

SPA CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI	0.20147	0.06891	63.0	464.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI	0.20212	0.06913	16.0	100.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI	0.20404	0.07376	25.0	100.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI	0.20413	0.07309	10.0	100.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, DIPENDENTI, PIL_PRO.CAPITE	0.20423	0.06627	10.0	215.0

Tabella 4.45: Tuning dei parametri del Random Forest: medie imprese Spa Centro

SPA SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, TOT_DISP.LIQUIDE, DIPENDENTI	1,096,268	0.93112	10.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP.LIQUIDE, DIPENDENTI, ISIC	1,100,740	0.90241	10.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, DIPENDENTI, ISIC	1,106,506	0.86790	63.0	215.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP.LIQUIDE, DIPENDENTI, ISIC	1,110,689	0.90269	16.0	1000.0
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP.LIQUIDE	1,115,375	0.93799	40.0	100.0

Tabella 4.46: Tuning dei parametri del Random Forest: medie imprese Spa Sud/Isole

SRL NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.71551	0.22560	10.0	464.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.71703	0.22486	16.0	464.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.71734	0.22521	25.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.71803	0.22602	10.0	1000.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.71856	0.22517	100.0	464.0

Tabella 4.47: Tuning dei parametri del Random Forest: medie imprese Srl Nord

SRL CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE	0.62976	0.17160	63.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE	0.63195	0.16252	25.0	100.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE	0.63384	0.16313	63.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE	0.63411	0.16476	16.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE	0.63535	0.16899	25.0	464.0

Tabella 4.48: Tuning dei parametri del Random Forest: medie imprese Srl Centro

SRL SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.51073	0.12547	40.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.51419	0.11496	16.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.51463	0.12224	40.0	464.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.51552	0.12040	16.0	464.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.51553	0.12374	40.0	1000.0

Tabella 4.49: Tuning dei parametri del Random Forest: medie imprese Srl Sud/Isole

SOCIETÀ COOPERATIVE NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP_LIQUIDE, DIPENDENTI	0.15648	0.07281	10.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DIPENDENTI, INFLAZIONE	0.15682	0.06785	16.0	215.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP_LIQUIDE, DIPENDENTI	0.15714	0.06725	40.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DIPENDENTI, CONSUMER.CONF.INDEX	0.15754	0.06910	100.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP_LIQUIDE, DIPENDENTI	0.15767	0.06931	100.0	1000.0

Tabella 4.50: Tuning dei parametri del Random Forest: medie imprese Società Cooperative Nord

SOCIETÀ COOPERATIVE CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
DEBITLOLTRE, DIPENDENTI	0.10336	0.10437	10.0	464.0
TOTALE.IMMOBILIZZAZIONI, DEBITLOLTRE, DIPENDENTI	0.10414	0.09049	16.0	100.0
DEBITLOLTRE, DIPENDENTI	0.10450	0.10593	10.0	100.0
TOTALE.IMMOBILIZZAZIONI, DEBITLOLTRE, DIPENDENTI, CONSUMER_CONF_INDEX	0.10472	0.09013	40.0	100.0
DEBITLOLTRE, DIPENDENTI	0.10507	0.10405	63.0	215.0

Tabella 4.51: Tuning dei parametri del Random Forest: medie imprese Società Cooperative Centro

SOCIETÀ COOPERATIVE SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITLOLTRE, DIPENDENTI	0.26978	0.18982	10.0	100.0
TOTALE.IMMOBILIZZAZIONI, CREDITLOLTRE, TOT_DISP_LIQUIDE, DEBITLOLTRE, DIPENDENTI	0.27441	0.20255	10.0	100.0
TOTALE.IMMOBILIZZAZIONI, CREDITLOLTRE, TOT_DISP_LIQUIDE, DEBITLOLTRE, DIPENDENTI	0.27534	0.19488	16.0	100.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITLOLTRE, DIPENDENTI	0.27540	0.18788	63.0	1000.0
TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITLOLTRE, DIPENDENTI	0.27615	0.18468	10.0	464.0

Tabella 4.52: Tuning dei parametri del Random Forest: medie imprese Società Cooperative Sud/Isole

GRANDI IMPRESE

SPA NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, DIPENDENTI	0.67510	0.12097	25.0	215.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, DIPENDENTI	0.67687	0.11808	16.0	215.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, DIPENDENTI	0.67706	0.11336	63.0	100.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, DIPENDENTI	0.67787	0.11729	40.0	215.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, DIPENDENTI	0.67790	0.11878	63.0	215.0

Tabella 4.53: Tuning dei parametri del Random Forest: grandi imprese Spa Nord

SPA CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
DIPENDENTI	0.38376	0.23198	63.0	100.0
DIPENDENTI	0.38595	0.23134	10.0	215.0
DIPENDENTI	0.38792	0.23074	16.0	100.0
DIPENDENTI	0.38857	0.22857	16.0	1000.0
DIPENDENTI	0.39039	0.22957	40.0	1000.0

Tabella 4.54: Tuning dei parametri del Random Forest: grandi imprese Spa Centro

SPA SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.59421	0.14795	40.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE, DIPENDENTI	0.59529	0.14013	10.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.59673	0.13557	16.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.59752	0.13895	25.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.59879	0.14696	10.0	464.0

Tabella 4.55: Tuning dei parametri del Random Forest: grandi imprese Spa Sud/Isole

SRL NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE	0.91042	0.18512	10.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, INDEX.PREZZI,CONSUMO	0.91135	0.16801	16.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, GRADO.APERTURA_COMMERCIALE	0.91878	0.17211	100.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE	0.92344	0.18528	16.0	464.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, INFLAZIONE	0.92365	0.17960	10.0	100.0

Tabella 4.56: Tuning dei parametri del Random Forest: grandi imprese Srl Nord

SRL CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI	0.26652	0.10461	40.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI	0.26697	0.11154	10.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI	0.26707	0.10916	100.0	464.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI	0.26800	0.10188	16.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI	0.26811	0.10728	16.0	1000.0

Tabella 4.57: Tuning dei parametri del Random Forest: grandi imprese Srl Centro

SRL SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE	0.60716	0.18204	10.0	215.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, GRADO.APERTURA.COMMERCIALE	0.60867	0.18479	40.0	100.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, GRADO.APERTURA.COMMERCIALE	0.60905	0.18489	10.0	464.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE	0.61083	0.18408	16.0	464.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE	0.61128	0.18283	10.0	464.0

Tabella 4.58: Tuning dei parametri del Random Forest: grandi imprese Srl Sud/Isole

SOCIETÀ COOPERATIVE NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE	0.15637	0.06423	63.0	215.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE	0.15716	0.06576	25.0	464.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE	0.15780	0.06528	63.0	464.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE	0.15783	0.06461	63.0	1000.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE	0.15789	0.06491	40.0	1000.0

Tabella 4.59: Tuning dei parametri del Random Forest: grandi imprese Società Cooperative Nord

Le restanti due casistiche, Società Cooperative Centro e Società Cooperative Sud/Isole, non contengono un numero di istanze tale da poter utilizzare l'operatore del *Loop Attributes Subset*.

4.6 Support Vector Regression

Il support vector regression estende l'algoritmo dell'SVM a problemi di regressione.

4.6.1 Funzionamento

In caso di problemi lineari di regressione si introduce la ϵ -insensitive loss function, che permette di ignorare errori minori della soglia ϵ , dove istanze con errore basso non contribuiscono all'ottimizzazione. Ciò rappresenta un'analogia con modelli prodotti dall'SVM. Essi, infatti, dipendono solo da un sottoinsieme dei dati di training, perché la funzione di costo per la costruzione

del modello non si preoccupa delle istanze che si trovano nel lato corretto del bordo decisionale. In regressione, questi punti sono quelli con errore minimo.

La funzione di regressione è definibile come

$$f(x) = w^T x + b \quad (4.2)$$

Analogamente all'SVM, la funzione $f(x)$ dovrebbe essere il più piatta possibile, minimizzando $\|w\|^2$. Al tempo stesso, è possibile introdurre delle variabili *slack*, ξ , ξ^* , variabili di rilassamento che consentono un margine di errore. Sulla base di questa idea, dato un training set $(t_1, y_1), (t_2, y_2), \dots, (t_n, y_n)$ dove y_i è continuo e $-\xi^* - \epsilon \leq y_i - y(t_i) \leq \epsilon + \xi_i$, il problema di ottimizzazione si pone nel seguente modo:

$$f(w) = \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i^* - \xi_i) \quad (4.3)$$

sotto vincoli

$$y_i \leq y(t_i) + \epsilon + \xi_i \quad (4.4)$$

$$y_i \geq y(t_i) - \xi_i^* - \epsilon \quad (4.5)$$

$$\xi_i^*, \xi_i \geq 0 \quad (4.6)$$

dove $C > 0$ è un parametro di regolarizzazione.

SVR non lineare

L'ipotesi della relazione lineare è spesso troppo semplicistica per caratterizzare problemi di regressione, motivo per cui si considera l'ipotesi di non linearità. L'idea di SVR per la regressione non lineare è quella di traslare lo spazio originale delle features in un nuovo spazio di caratteristiche dimensionali superiori, $\phi : \mathbb{R}^n \rightarrow \chi$, dove χ è lo spazio delle features, attraverso le funzioni di Kernel. Quest'ultime dipendono solamente dalle variabili originali e fanno in modo che la relazione non lineare possa essere approssimata con le stesse modalità che si avrebbero nel caso lineare.

È possibile considerare diverse funzioni di Kernel nella costruzione di un modello SVR, motivo per cui il tipo di Kernel diventa di fatto un iper-parametro da ottimizzare attraverso procedure di *cross-validation*.

4.6.2 Analisi

Per ogni casistica in esame si è proceduto con l'ottimizzazione del parametro *Kernel type*, per individuare la tipologia che meglio si adattasse ad ogni caso. Così come svolto in precedenza con l'utilizzo di *Decision Tree* e *Random Forest*, anche per l'SVR si è optato per individuare la combinazione di variabili e relativi parametri con associato il minimo errore.

In particolare, le tipologie di funzioni di Kernel prese in considerazione sono riportate in tabella 4.60:

KERNEL FUNCTION	FORMULA
Lineare	$K(x, y) = x^T y$
Polinomiale	$K(x, y) = (\gamma x^T y + r)^d, \gamma > 0$
Sigmoidale	$K(x, y) = \exp(-\gamma \ x - y\ ^2), \gamma > 0$
Radial Basis Function	$K(x, y) = \tanh(\gamma x^T y + r)$

Tabella 4.60: Tipologie di funzioni di Kernel

Dai risultati del *Loop Attributes Subset*, sono state individuate le cinque migliori combinazioni con il relativo tipo di Kernel, al fine di procedere con un'ulteriore ottimizzazione dei parametri propri della specifica funzione di Kernel. Oltre al parametro di regolarizzazione C comune a tutte le funzioni, i parametri variano da funzione a funzione. Le tabelle 4.61 e 4.62 mostrano rispettivamente i parametri caratteristici di ogni funzione e il relativo settaggio nel *tuning*.

KERNEL FUNCTION	PARAMETRI
Lineare	C
Polinomiale	$C, \gamma, \text{grado del polinomio}$
Sigmoidale	C, γ
Radial Basis Function	C, γ

Tabella 4.61: Parametri specifici delle funzioni di Kernel

PARAMETRO	TIPOLOGIA DI SETTAGGIO	SETTAGGIO
C	lista	0,10,100,1000,10000
γ	lista	0.00001, 0.0001, 0.001, 0.01, 0
Grado del polinomio	range	2-5 linear

Tabella 4.62: Setting dei parametri delle funzioni di Kernel

Di seguito, per ogni casistica, sono riportate le combinazioni di features e relativi parametri per ognuna delle precedenti cinque miglior combinazioni con associata la propria funzione di Kernel.

PICCOLE IMPRESE

SPA NORD

FEATURES COMBINATION	RMSE	SD RMSE	MAXIMAL DEPTH	NUMBER OF TREES
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE	0.7451922400236977	0.13332449111390138	lineare	10.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, PIL_PRO_CAPITE	0.74549	0.13412	lineare	10000.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.74601	0.13375	lineare	10.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, INFLAZIONE	0.74618	0.13337	lineare	10.0
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI_A_BREVE, TOT_DISP_LIQUIDE	0.74658	0.13357	lineare	10000.0

Tabella 4.63: Tuning dei parametri del SVR: piccole imprese Spa Nord

SPA CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, CREDITI A BREVE, CREDITI OLTRE, TOT DISP LIQUIDE	1.58890	2.63133	lineare	10.0	
VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, TOT DISP LIQUIDE, PIL PRO CAPITE	1.57429	2.61413	lineare	10.0	
VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, TOT DISP LIQUIDE	1.56954	2.61990	lineare	10.0	
VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, CREDITI A BREVE, TOT DISP LIQUIDE, PIL PRO CAPITE, INFLAZIONE, ISIC	1.59056	2.61041	sigmoide	100000.0	0.0001
VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, TOT DISP LIQUIDE, ISIC	1.57500	2.61488	lineare	10.0	

Tabella 4.64: Tuning dei parametri del SVR: piccole imprese Spa Centro

SPA SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
COSTI PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, DIPENDENTI	0.63348	0.10784	lineare	1000.0	
COSTI PRODUZIONE, TOTALE IMMOBILIZZAZIONI, DIPENDENTI	0.63392	0.10982	lineare	1000.0	
COSTI PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, TOT DISP LIQUIDE, DIPENDENTI	0.63474	0.10268	lineare	0.0	
COSTI PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, TOT DISP LIQUIDE, DIPENDENTI, PIL PRO CAPITE	0.63494	0.10333	sigmoide	1000.0	0.001
COSTI PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI OLTRE, TOT DISP LIQUIDE, DIPENDENTI, ISIC	0.63570	0.10456	sigmoide	100000.0	0.00001

Tabella 4.65: Tuning dei parametri del SVR: piccole imprese Spa Sud/Isole

SRL NORD

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, INDEX_PREZZI_CONSUMO, BUSINESS_CONF_INDEX	0.93535	0.22269	lineare	10000.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, INDEX_PREZZI_CONSUMO	0.93562	0.22261	lineare	0.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI_OLTRE, TOT_DISP_LIQUIDE, INDEX_PREZZI_CONSUMO	0.93563	0.22229	lineare	0.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, BUSINESS_CONF_INDEX	0.935671	0.22305	lineare	10000.0
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI_OLTRE, TOT_DISP_LIQUIDE, INDEX_PREZZI_CONSUMO, BUSINESS_CONF_INDEX	0.935730	0.22218	lineare	0.0

Tabella 4.66: Tuning dei parametri del SVR: piccole imprese Srl Nord

SRL CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C
VALORE DELLA PRODUZIONE, CREDITI_OLTRE, TOT_DISP_LIQUIDE, DEBITI_OLTRE, INDEX_PREZZI_CONSUMO	0.90951	0.24441	lineare	0.0
VALORE DELLA PRODUZIONE, CREDITI_OLTRE, TOT_DISP_LIQUIDE, DEBITI_OLTRE, INDEX_PREZZI_CONSUMO, INFLAZIONE	0.90973	0.24452	lineare	0.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DEBITI_OLTRE, INDEX_PREZZI_CONSUMO, INFLAZIONE	0.90991	0.24782	lineare	100.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DEBITI_OLTRE, INDEX_PREZZI_CONSUMO	0.91001	0.24763	lineare	10.0
VALORE DELLA PRODUZIONE, TOT_DISP_LIQUIDE, DEBITI_OLTRE, INDEX_PREZZI_CONSUMO, BUSINESS_CONF_INDEX	0.91012	0.24756	lineare	100.0

Tabella 4.67: Tuning dei parametri del SVR: piccole imprese Srl Centro

SRL SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA	GRADO
VALORE.DELLA.PRODUZIONE, TOT.DISP.LIQUIDE, DEBITLOLTRE, INDEX.PREZZI.CONSUMO, INFLAZIONE, ISIC	0.69989	0.04194	polinomiale	1000.0	0.01	3.0
VALORE.DELLA.PRODUZIONE, TOT.DISP.LIQUIDE, DEBITLOLTRE, PIL.PRO.CAPITE, INDEX.PREZZI.CONSUMO, BUSINESS.CONF_INDEX, ISIC	0.70547	0.06094	polinomiale	1000.0	0.01	3.0
VALORE.DELLA.PRODUZIONE, TOT.DISP.LIQUIDE, DEBITLOLTRE, PIL.PRO.CAPITE, INDEX.PREZZI.CONSUMO, INFLAZIONE, ISIC	0.71147	0.06068	polinomiale	1000.0	0.01	3.0
VALORE.DELLA.PRODUZIONE, TOT.DISP.LIQUIDE, DEBITLOLTRE, PIL.PRO.CAPITE, INDEX.PREZZI.CONSUMO, BUSINESS.CONF_INDEX, INFLAZIONE, ISIC	0.71268	0.06993	polinomiale	100.0	0.01	3.0
VALORE.DELLA.PRODUZIONE, TOT.DISP.LIQUIDE, DEBITLOLTRE, PIL.PRO.CAPITE, INFLAZIONE, ISIC	0.72935	0.08506	polinomiale	1000.0	0.01	3.0

Tabella 4.68: Tuning dei parametri del SVR: piccole imprese Srl Sud/Isole

SOCIETÀ COOPERATIVE NORD

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA	GRADO
TOTALE.IMMOBILIZZAZIONI, TOT.DISP.LIQUIDE, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.25914	0.10832	polinomiale	100.0	0.1	3.0
TOTALE.IMMOBILIZZAZIONI, TOT.DISP.LIQUIDE, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.25914	0.10832	polinomiale	100.0	0.1	3.0
TOTALE.IMMOBILIZZAZIONI, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.25952	0.10861	polinomiale	10000.0	0.1	3.0
TOTALE.IMMOBILIZZAZIONI, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.25952	0.10861	polinomiale	10000.0	0.1	3.0
TOTALE.IMMOBILIZZAZIONI, INDEX.PREZZI.CONSUMO	0.26673	0.11079	polinomiale	10000.0	0.1	3.0

Tabella 4.69: Tuning dei parametri del SVR: piccole imprese Società Cooperative Nord

SOCIETÀ COOPERATIVE CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITI.OLTRE, TOT_DISP.LIQUIDE, DEBITI.OLTRE, DIPENDENTI, PIL.PRO.CAPITE, INDEX.PREZZI.CONSUMO	0.15639	0.12758	rbf	1000.0	0.001
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITI.OLTRE, TOT_DISP.LIQUIDE, DEBITI.OLTRE, DIPENDENTI, INDEX.PREZZI.CONSUMO, CONSUMER.CONF.INDEX	0.15811	0.12776	rbf	1000.0	0.001
TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP.LIQUIDE, DEBITI.OLTRE, DIPENDENTI, INDEX.PREZZI.CONSUMO, CONSUMER.CONF.INDEX	0.16001	0.11892	rbf	10.0	0.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP.LIQUIDE, DEBITI.OLTRE, DIPENDENTI, PIL.PRO.CAPITE, INDEX.PREZZI.CONSUMO	0.16121	0.12802	rbf	1000.0	0.001
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP.LIQUIDE, DEBITI.OLTRE, DIPENDENTI, INDEX.PREZZI.CONSUMO, CONSUMER.CONF.INDEX	0.16137	0.12916	rbf	1000.0	0.001

Tabella 4.70: Tuning dei parametri del SVR: piccole imprese Società Cooperative Centro

SOCIETÀ COOPERATIVE SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA	GRADO
TOTALE.IMMOBILIZZAZIONI, INDEX.PREZZI.CONSUMO	0.43417	0.18242	polinomiale	100.0	0.1	3.0
TOTALE.IMMOBILIZZAZIONI, INDEX.PREZZI.CONSUMO	0.43417	0.18242	polinomiale	100.0	0.1	3.0
TOTALE.IMMOBILIZZAZIONI, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.44219	0.17896	polinomiale	100.0	0.1	3.0
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DIPENDENTI, PIL.PRO.CAPITE	0.44297	0.18226	sigmoide	10.0	0.01	
TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DIPENDENTI, PIL.PRO.CAPITE, CONSUMER.CONF.INDEX, INFLAZIONE	0.44815	0.18656	sigmoide	100.0	0.001	

Tabella 4.71: Tuning dei parametri del SVR: piccole imprese Società Cooperative Sud/Isole

MEDIE IMPRESE

SPA NORD

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA	GRADO
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT.DISP.LIQUIDE, GRADO.APERTURA.COMMERCIALE, INFLAZIONE	0.91204	0.13351	polinomiale	10.0	0.1	4.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT.DISP.LIQUIDE, DEBITI.OLTRE, GRADO.APERTURA.COMMERCIALE, INFLAZIONE	0.95074	0.26457	polinomiale	10000.0	0.01	4.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT.DISP.LIQUIDE, DEBITI.OLTRE, GRADO.APERTURA.COMMERCIALE	0.96965	0.25009	polinomiale	10.0	0.1	4.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT.DISP.LIQUIDE, INFLAZIONE	1.05581	0.21888	polinomiale	10.0	0.1	4.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT.DISP.LIQUIDE, INDEX.PREZZI.CONSUMO, INFLAZIONE	1.06378	0.30823	polinomiale	10.0	0.1	3.0

Tabella 4.72: Tuning dei parametri del SVR: medie imprese Spa Nord

SPA CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA	GRADO
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, CREDITI.OLTRE, DIPENDENTI, GRADO.APERTURA.COMMERCIALE	0.29114	0.09559	polinomiale	10000.0	0.0	4.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, CREDITI.OLTRE, DIPENDENTI, INDEX.PREZZI.CONSUMO, GRADO.APERTURA.COMMERCIALE	0.29840	0.07660	polinomiale	10000.0	0.0	3.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, DIPENDENTI, GRADO.APERTURA.COMMERCIALE	0.30233	0.10432	polinomiale	10.0	0.1	4.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, DIPENDENTI, GRADO.APERTURA.COMMERCIALE	0.31746	0.07801	polinomiale	100.0	0.0	2.0
TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, DIPENDENTI, INDEX.PREZZI.CONSUMO, GRADO.APERTURA.COMMERCIALE	0.32348	0.10698	polinomiale	0.0	0.001	3.0

Tabella 4.73: Tuning dei parametri del SVR: medie imprese Spa Centro

SPA SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA	GRADO
CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.93902	0.33251	polinomiale	0.0	0.0	3.0
VALORE.DELLA.PRODUZIONE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.95916	0.33287	polinomiale	0.0	1.0E-5	3.0
VALORE.DELLA.PRODUZIONE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, DIPENDENTI, INDEX.PREZZI.CONSUMO	0.95997	0.33189	polinomiale	0.0	1.0E-4	3.0
CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, DIPENDENTI, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.96080	0.33118	polinomiale	0.0	1.0E-5	3.0
CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, DIPENDENTI, INDEX.PREZZI.CONSUMO	0.96191	0.32979	polinomiale	0.0	1.0E-5	3.0

Tabella 4.74: Tuning dei parametri del SVR: medie imprese Spa Sud/Isole

SRL NORD

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI, INDEX.PREZZI.CONSUMO, INFLAZION	0.80039	0.22692	lineare	10000.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, INFLAZIONE, ISIC	0.80116	0.23184	lineare	0.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI, INFLAZIONE	0.80118	0.22706	lineare	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, INFLAZIONE	0.80120	0.23206	lineare	0.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI, INFLAZIONE, ISIC	0.80121	0.22751	lineare	0.0

Tabella 4.75: Tuning dei parametri del SVR: medie imprese Srl Nord

SRL CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITLA.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, INDEX.PREZZI.CONSUMO	0.54626	0.17378	sigmoide	100.0	0.01
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITLA.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, INDEX.PREZZI.CONSUMO, ISIC	0.55031	0.17523	sigmoide	100.0	0.01
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITLA.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, INDEX.PREZZI.CONSUMO	0.55872	0.17848	sigmoide	100.0	0.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, INDEX.PREZZI.CONSUMO	0.58176	0.16235	sigmoide	100.0	0.01
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, INDEX.PREZZI.CONSUMO	0.58429	0.16579	lineare	10.0	

Tabella 4.76: Tuning dei parametri del SVR: medie imprese Srl Centro

SRL SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, INDEX.PREZZI.CONSUMO	0.53434	0.12340	lineare	10000.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE, INDEX.PREZZI.CONSUMO	0.53467	0.12013	lineare	0.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DIPENDENTI, INDEX.PREZZI.CONSUMO	0.53490	0.12207	lineare	10.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, CREDITI.OLTRE, TOT_DISP_LIQUIDE, INDEX.PREZZI.CONSUMO	0.53434	0.12340	lineare	10000.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE, INDEX.PREZZI.CONSUMO, ISIC	0.53530	0.12165	lineare	0.0

Tabella 4.77: Tuning dei parametri del SVR: medie imprese Srl Sud/Isole

SOCIETÀ COOPERATIVE NORD

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DEBITI.OLTRE, DIPENDENTI, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.13719	0.04687	sigmoide	10.0	0.0
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE, INDEX.PREZZI.CONSUMO	0.13766	0.05288	sigmoide	0.0	0.01
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DEBITI.OLTRE, DIPENDENTI, INFLAZIONE	0.13810	0.04725	sigmoide	1000.0	0.001
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DEBITI.OLTRE, DIPENDENTI, INFLAZIONE	0.14068	0.05448	sigmoide	10.0	0.01
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, DEBITI.OLTRE, DIPENDENTI, INDEX.PREZZI.CONSUMO, INFLAZIONE	0.14188	0.05282	sigmoide	1000.0	0.001

Tabella 4.78: Tuning dei parametri del SVR: medie imprese Società Cooperative Nord

SOCIETÀ COOPERATIVE CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITLA.BREVE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, DIPENDENTI	0.08308	0.05306	sigmoide	0.0	0.0
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, CREDITLA.BREVE, DEBITI.OLTRE, DIPENDENTI	0.08319	0.05512	sigmoide	0.0	0.0
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, DIPENDENTI	0.08366	0.05864	sigmoide	0.0	0.0
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOTALE.RIMANENZE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, DIPENDENTI	0.08493	0.05481	lineare	10.0	
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE, DIPENDENTI, INDEX.PREZZI.CONSUMO	0.08657	0.05756	lineare	0.0	

Tabella 4.79: Tuning dei parametri del SVR: medie imprese Società Cooperative Centro

SOCIETÀ COOPERATIVE SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA	GRADO
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, CREDITLA.BREVE, TOT_DISP_LIQUIDE, DIPENDENTI	0.29474	0.24994	sigmoide	100.0	0.0	
COSTI.PRODUZIONE, CREDITLA.BREVE, TOT_DISP_LIQUIDE, DIPENDENTI	0.306087	0.21688	rbf	100.0	0.0	
COSTI.PRODUZIONE, TOT_DISP_LIQUIDE, INDEX_PREZZI_CONSUMO	0.33540	0.23474	polinomiale	100.0	0.1	3.0
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI, INFLAZIONE	0.34995	0.28571	sigmoide	0.0	0.0	
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI, INFLAZIONE	0.35459	0.28428	sigmoide	0.0	0.0	

Tabella 4.80: Tuning dei parametri del SVR: medie imprese Società Cooperative Sud/Isole

GRANDI IMPRESE

SPA NORD

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.59687	0.06796	lineare	10000.0	
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, DIPENDENTI	0.59878	0.07002	lineare	10000.0	
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITLA.BREVE, TOT_DISP_LIQUIDE, DIPENDENTI	0.59801	0.06798	lineare	10000.0	
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	0.60056	0.06811	sigmoide	10000.0	0.0001
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, DIPENDENTI	0.60361	0.06926	sigmoide	1000.0	0.001

Tabella 4.81: Tuning dei parametri del SVR: grandi imprese Spa Nord

SPA CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, TOT.DISP.LIQUIDE, DIPENDENTI	0.42670	0.18174	rbf	10.0	0.1
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, TOT.DISP.LIQUIDE, DIPENDENTI	0.45226	0.19378	rbf	10.0	0.1
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, CREDITI.OLTRE, TOT.DISP.LIQUIDE, DIPENDENTI	0.47310	0.18961	rbf	10.0	0.1
COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, CREDITI.OLTRE, TOT.DISP.LIQUIDE, DIPENDENTI	0.48929	0.19036	rbf	10.0	0.1
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.A.BREVE, CREDITI.OLTRE, TOT.DISP.LIQUIDE	0.52160	0.22507	rbf	100.0	0.01

Tabella 4.82: Tuning dei parametri del SVR: grandi imprese Spa Centro

SPA SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT.DISP.LIQUIDE, DIPENDENTI, INFLAZIONE	0.6413766280309015	0.1878891922086651	linear	100.0	
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOT.DISP.LIQUIDE, DIPENDENTI, INFLAZIONE	0.64696	0.16201	sigmoid	10000.0	0.0001
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOT.DISP.LIQUIDE, DIPENDENTI	0.65003	0.16632	sigmoid	10000.0	0.0002
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOT.DISP.LIQUIDE, DIPENDENTI	0.65233	0.17052	linear	10.0	
VALORE.DELLA.PRODUZIONE, COSTI.PRODUZIONE, TOT.DISP.LIQUIDE, DIPENDENTI, INFLAZIONE	0.65283	0.17349	linear	10000.0	

Tabella 4.83: Tuning dei parametri del SVR: grandi imprese Spa Sud/Isole

SRL NORD

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, INDEX_PREZZI.CONSUMO	114740	0.35865	linear	1000.0	
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, BUSINESS_CONF_INDEX	114730	0.35825	linear	100.0	
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	114829	0.35973	linear	100.0	
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI.OLTRE	115162	0.35574	linear	100.0	
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	114944	0.35984	linear	10000.0	0.0001

Tabella 4.84: Tuning dei parametri del SVR: grandi imprese Srl Nord

SRL CENTRO

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE, ISIC	0.27851	0.17472	sigmoide	10000.0	0.01
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.29089	0.14440	sigmoide	1000.0	0.01
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, DEBITI.OLTRE	0.30797	0.14103	sigmoide	1000.0	0.001
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, DEBITI.OLTRE	0.31198	0.13281	lineare	0.0	
VALORE.DELLA.PRODUZIONE, TOTALE.IMMOBILIZZAZIONI, CREDITI.OLTRE, TOT_DISP_LIQUIDE, DEBITI.OLTRE	0.31198	0.13281	lineare	0.0	

Tabella 4.85: Tuning dei parametri del SVR: grandi imprese Srl Centro

SRL SUD/ISOLE

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, PIL_PRO_CAPITE, BUSINESS_CONF_INDEX, INFLAZIONE, ISIC	0.76795	0.205477	sigmoide	10000.0	0.0001
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, PIL_PRO_CAPITE, ISIC	0.76975	0.20725	sigmoide	100.0	
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, PIL_PRO_CAPITE, ISIC	0.77444	0.20758	sigmoide	10000.0	0.001
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, PIL_PRO_CAPITE, INFLAZIONE, ISIC	0.77500	0.18249	lineare	10.0	
VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, PIL_PRO_CAPITE, INDEX_PREZZI_CONSUMO, ISIC	0.77771	0.20478	lineare	10000.0	

Tabella 4.86: Tuning dei parametri del SVR: grandi imprese Srl Sud/Isole

SOCIETÀ COOPERATIVE NORD

FEATURES COMBINATION	RMSE	SD RMSE	KERNEL	C	GAMMA	GRADO
TOTALE IMMOBILIZZAZIONI, TOTALE RIMANENZE, TOT_DISP_LIQUIDE	0.14462	0.04610	polinomiale	1000.0	0.1	3.0
TOTALE IMMOBILIZZAZIONI, TOTALE RIMANENZE, TOT_DISP_LIQUIDE, DEBITI_OLTRE	0.15343	0.05569	polinomiale	10000.0	0.0	3.0
TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DEBITI_OLTRE	0.16460	0.07628	polinomiale	1000.0	0.1	3.0
TOTALE IMMOBILIZZAZIONI, DEBITI_OLTRE	0.17817	0.10192	polinomiale	0.0	0.0	3.0
TOTALE IMMOBILIZZAZIONI, TOTALE RIMANENZE, TOT_DISP_LIQUIDE, INDEX_PREZZI_CONSUMO	0.18308	0.14065	polinomiale	10000.0	0.01	3.0

Tabella 4.87: Tuning dei parametri del SVR: grandi imprese Società Cooperative Nord

Le restanti due casistiche, Società Cooperative Centro e Società Cooperative Sud/Isole, non contengono un numero di istanze tale da poter utilizzare l'operatore del *Loop Attribute Subset*.

Capitolo 4

5.1 Risultati

A seguito del processo esposto nel Capitolo 3, si è proceduto ad individuare, per ogni cluster, la migliore combinazione di variabili e parametri di ogni algoritmo utilizzato. Quest'ultime sono state utilizzate per validare in test i modelli ottenuti dall'uso di *Decision Tree*, *Random Forest* e *SVR*. Dal confronto dei rispettivi RMSE, è stato identificato, in ogni casistica, l'algoritmo con le performance più elevate. Le Tabelle 5.1-5.3 mostrano i risultati del confronto, riportando i valori dell'RMSE in *training*, della *Standard Deviation* dell'RMSE in *training* e dell'RMSE in *test*, nonché gli stessi opportunamente de-normalizzati. Al fine di poter effettuare una valutazione della bontà dei diversi modelli, è stato inserito il *range* della variabile dipendente *Ebitda Margin*, relativo ad ogni cluster, in riferimento ai de-normalizzati.

5.1.1 Piccole Imprese

FORMA GIURIDICA	AREA	ALGORITMO	FEATURES	PARAMETRI	RMSE TRAINING	SD RMSE TRAINING	RMSE TEST	SD RMSE TEST	RANGE ERITDA MARGIN DE-NORMALIZZATO	RMSE TRAINING DE-NORMALIZZATO	SD RMSE TRAINING DE-NORMALIZZATO	RMSE TEST DE-NORMALIZZATO
SFA	Need	RANDOM FOREST	VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI A BREVE, CREDITI LOUTRE, TOT DISP LIQUIDE,	MAX DEPTH= 100 N. TREES=215	0.6664	0.1228	0.978	0.1226	[-0.483 ; 0.333]	0.6825	0.1173	0.1473
SFA	Contro	RANDOM FOREST	VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, CREDITI A BREVE, DIPENDENTI	MAX DEPTH= 25 N. TREES=100	1.465	2.463	0.7581	0.1843	[-2.221 ; 0.305]	0.2904	0.1306	0.1306
SFA	Sm/Isok	RANDOM FOREST	COSTI PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI A BREVE, DIPENDENTI	MAX DEPTH= 16 N. TREES=464	0.4579	0.1288	0.593	0.1062	[-0.032 ; 0.232]	0.6827	0.1180	0.1180
SRL	Need	RANDOM FOREST	VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT DISP LIQUIDE, DEBITI LOUTRE	MAX DEPTH= 100 N. TREES=100	0.9009	0.2308	0.923	0.1431	[-1.318 ; 1.114]	0.9009	0.1431	0.1431
SRL	Contro	RANDOM FOREST	VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI LOUTRE, TOT DISP LIQUIDE, DEBITI LOUTRE	MAX DEPTH= 16 N. TREES=100	0.8034	0.2798	1.005	0.1340	[-0.422 ; 1]	0.6942	0.1493	0.1493
SRL	Sm/Isok	RANDOM FOREST	VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI LOUTRE, TOT DISP LIQUIDE, DEBITI LOUTRE	MAX DEPTH= 100 N. TREES=100	0.6288	0.1291	0.735	0.1307	[-0.119 ; 0.884]	0.6288	0.1288	0.1288
Società Cooperative	Need	RANDOM FOREST	TOTALE IMMOBILIZZAZIONI, CREDITI LOUTRE, TOT DISP LIQUIDE, DIPENDENTI	MAX DEPTH= 100 N. TREES=215	0.2318	0.0986	0.266	0.0906	[-0.068 ; 0.346]	0.0834	0.0932	0.0932
Società Cooperative	Contro	RANDOM FOREST	TOTALE IMMOBILIZZAZIONI, CREDITI LOUTRE, DEBITI LOUTRE, DIPENDENTI, INDEX PREZZI CONSUMO	MAX DEPTH= 25 N. TREES= 100	0.1422	0.1376	0.139	0.0538	[-0.01 ; 0.09]	0.0834	0.0835	0.0835
Società Cooperative	Sm/Isok	RANDOM FOREST	TOTALE IMMOBILIZZAZIONI, TOTALE RIMANENZE, TOT DISP LIQUIDE, DEBITI LOUTRE	MAX DEPTH= 63 N. TREES=100	0.3128	0.1173	0.34	0.0967	[-0.086 ; 0.241]	0.0819	0.0988	0.0988

Tabella 5.1: Risultati finali Piccole Imprese

5.1.2 Medie Imprese

FORMA GIURIDICA	AREA	ALGORITMO	FEATURES	PARAMETRI	RMSE TRAINING	SD RMSE TRAINING	RMSE TEST	RANGE EBITDA MARGIN DE-NORMALIZZATO	RMSE TRAINING DE-NORMALIZZATO	SD RMSE TRAINING DE-NORMALIZZATO	RMSE TEST DE-NORMALIZZATO
SPA	Nord	RANDOM FOREST	TOTALE,IMMOBILIZZAZIONI, TOTL,DISP,LIQUIDE, DIPENDENTI	MAX DEPUT= 100 N_TREES=100	0,8794	0,2305	0,884	[0,554 ; 0,879]	0,1398	0,0905	0,1401
SPA	Centro	RANDOM FOREST	TOTALE,IMMOBILIZZAZIONI, CREDITI,OUTRE, DIPENDENTI	MAX DEPUT= 60 N_TREES=464	0,2014	0,6689	0,554	[0,027 ; 0,196]	0,6883	0,0782	0,1151
SPA	Staf/boab	RANDOM FOREST	TOTALE,IMMOBILIZZAZIONI, CREDITI, A,BREVE, TOTL,DISP,LIQUIDE, DIPENDENTI	MAX DEPUT= 10 N_TREES=100	1,0982	0,9311	0,756	[-0,142 ; 1,004]	0,1563	0,1437	0,1304
SRL	Nord	RANDOM FOREST	VALORE,DELLA,PRODUZIONE, TOTALE,IMMOBILIZZAZIONI, TOTL,DISP,LIQUIDE, DEBITI,OUTRE	MAX DEPUT= 10 N_TREES=464	0,7155	0,2256	0,938	[0,153 ; 0,489]	0,1273	0,0901	0,1427
SRL	Centro	RANDOM FOREST	VALORE,DELLA,PRODUZIONE, TOTALE,IMMOBILIZZAZIONI, CREDITI,OUTRE, TOTL,DISP,LIQUIDE	MAX DEPUT= 63 N_TREES=100	0,6297	0,1716	0,627	[0,055 ; 0,281]	0,1268	0,0860	0,1206
SRL	Staf/boab	SVR	VALORE,DELLA,PRODUZIONE, TOTALE,IMMOBILIZZAZIONI, CREDITI, A,BREVE, CREDITI,OUTRE, TOTL,DISP,LIQUIDE, INVESt,PREZZI,CONSUMO	KERNEL=LINEARE C=10000	0,5343	0,1241	0,888	[0,341 ; 0,261]	0,1136	0,0523	0,1389
Società Cooperative	Nord	RANDOM FOREST	TOTALE,IMMOBILIZZAZIONI, TOTALE,IRMANENZE, TOTL,DISP,LIQUIDE, DIPENDENTI	MAX DEPUT= 10 N_TREES=100	0,1564	0,0728	0,173	[0,001 ; 0,162]	0,0848	0,0785	0,0861
Società Cooperative	Centro	SVR	COSTI,PRODUZIONE, TOTALE,IMMOBILIZZAZIONI, TOTALE,IRMANENZE, CREDITI, A,BREVE, TOTL,DISP,LIQUIDE, DEBITI,OUTRE, DIPENDENTI	C=4 GAMMA=0	0,0830	0,0330	0,111	[0,001 ; 0,124]	0,0793	0,0770	0,0814
Società Cooperative	Staf/boab	SVR	VALORE,DELLA,PRODUZIONE, COSTI,PRODUZIONE, CREDITI, A,BREVE, TOTL,DISP,LIQUIDE, DIPENDENTI	KERNEL=RF C=100 GAMMA=0	0,2007	0,2409	0,217	[0,001 ; 0,154]	0,0934	0,0919	0,0894

Tabella 5.2: Risultati finali Medie Imprese

5.1.3 Grandi Imprese

FORMA GIURIDICA	AREA	ALGORITMO	FEATURES	PARAMETRI	RMSE TRAINING	SD RMSE TRAINING	RANGE EBITDA MARGIN DE-NORMALIZZATO	RMSE TRAINING DE-NORMALIZZATO	SD RMSE TRAINING DE-NORMALIZZATO	RMSE TEST DE-NORMALIZZATO
SFA	Nord	SVR	VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	KERNEL=LINEARE C=10000	0.5968	0.0679	[-0.21; 0.27]	0.1183	0.0781	0.1301
SFA	Centro	SVR	VALORE DELLA PRODUZIONE, COSTI PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI A BREVE, TOT_DISP_LIQUIDE, DIPENDENTI	KERNEL=BBF C=10 GAMMA=0.1	0.4287	0.1817	[-0.083; 0.251]	0.1054	0.0868	0.1383
SFA	Staf/Isola	RANDOM FOREST	VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE, DIPENDENTI	MAX DEPTH=40 N. TREES=100	0.5942	0.1179	[-0.092; 0.223]	0.1181	0.0842	0.1198
SRL	Nord	RANDOM FOREST	VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI A BREVE, TOT_DISP_LIQUIDE	MAX DEPTH=10 N. TREES=215	0.9104	0.1851	[-0.181; 0.757]	0.1421	0.0870	0.1410
SRL	Centro	SVR	VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, CREDITI A BREVE, TOT_DISP_LIQUIDE, DEBITI OLTRE I 50C	KERNEL=SIGMOIDE C=10000 GAMMA=0.01	0.2785	0.1747	[0.01; 0.212]	0.0941	0.0862	0.1194
SRL	Staf/Isola	RANDOM FOREST	VALORE DELLA PRODUZIONE, TOTALE IMMOBILIZZAZIONI, TOT_DISP_LIQUIDE	MAX DEPTH=10 N. TREES=215	0.6071	0.1820	[0.02; 0.291]	0.1191	0.0868	0.1323
Società Cooperative	Nord	SVR	TOTALE IMMOBILIZZAZIONI, TOTALE RIMANENZE, TOT_DISP_LIQUIDE	KERNEL=POLINOMIALE C=1000 GAMMA=0.1 GRADO=3	0.1446	0.0491	[0.04; 0.101]	0.0889	0.0765	0.0884

Tabella 5.3: Risultati finali Grandi Imprese

5.1.4 Discussione

Nella valutazione della bontà del modello di regressione attraverso la metrica dell'RMSE, l'interpretazione dell'errore è commisurata al range di valori che assume la variabile dipendente. Considerando tutte le istanze del dataset, il range dell'*Ebitda Margin* è di $[-2.321 ; 1.114]$ con media di 0.073 e *Standard Deviation* del 0.076. Così come mostra la Figura 5.1, la maggiorparte dei valori si concentrano all'interno dell'intervallo $[-0.25 ; 0.5]$, che risulta essere piuttosto ristretto.

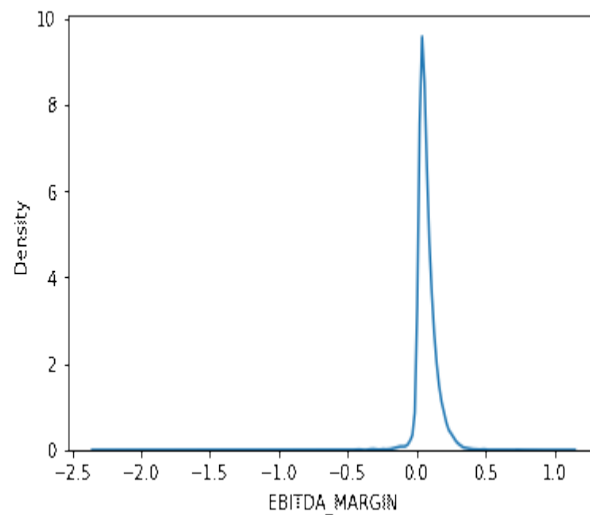


Figura 5.1: Range Ebitda Margin

I modelli proposti non performano bene nelle casistiche esaminate.

Considerando in particolare le grandi imprese, uno dei fattori che incide sulla scarsa capacità predittiva dei modelli è l'assenza di ulteriori variabili esogene di

tipo macroeconomico. Poiché la dimensionalità elevata di un'impresa identifica la presenza di questa all'interno di più stati e/o continenti, alcune dinamiche di mercato come politiche di economia monetaria (ad esempio variazioni del tasso di cambio), differente capacità di acquisto, quadro normativo ed ulteriori variabili micro e macroeconomiche dei paesi esteri, influenzano gli aspetti legati alla redditività d'impresa. Una logica di questo tipo rafforza ulteriormente il presupposto, già parzialmente esaminato nel corso dei capitoli, secondo cui l'impresa non opera come identità a sé stante ma, al contrario, in continuo rapporto con l'ambiente di appartenenza. In quest'ottica, un sistema completo ed integrato di indici capaci di rappresentare l'ambiente esterno, rafforza la capacità di valutare l'assetto economico di un'azienda.

In aggiunta, con riferimento anche alle piccole e medie imprese, alcune variabili strutturali, strategie manageriali, modelli di business adottati, posizionamento del comparto, nonché la maggiore o minore forza contrattuale di forze competitive (come clienti e fornitori), possono avere particolare impatto anche sull'andamento della redditività aziendale.

I fattori critici di successo, intesi come variabili chiavi per lo sviluppo di un'organizzazione aziendale, fanno riferimento, ad esempio, all'importanza del brand, ad adeguate politiche di marketing, al *know how* in possesso dell'azienda, alla qualità dei processi. Essi rappresentano il patrimonio intangibile dell'impresa e, seppur molto importanti, spesso non sono quantificabili.

li, nonché più difficilmente reperibili, e quindi non valutabili all'interno dei modelli precedentemente considerati.

Inoltre, l'elevata segmentazione delle istanze del dataset è, da un lato, propedeutica all'individuazione di gruppi il più possibile omogenei al loro interno, dall'altro riduce la numerosità campionaria di ciascuna casistica. Ciò potrebbe incidere sulla bontà dei modelli e, in quest'ottica, si potrebbe pensare a limitare la suddivisione soltanto su base dimensionale e territoriale.

Analisi delle variabili

Le variabili che più ricorrono nei risultati delle varie casistiche sono

- Valore della produzione
- Costi della produzione
- Totale immobilizzazioni
- Totale disponibilità liquide

Il valore della produzione e i costi della produzione sono le due voci di bilancio che determinano l'indice dell'*Ebitda Margin*, motivo per cui è molto verosimile che siano considerate tra quelle più rappresentative del fenomeno preso in esame.

Il totale delle immobilizzazioni, che al loro interno comprendono sia quelle di natura tangibile e intangibile, sono formate perlopiù da immobilizzazioni

materiali. Se un'impresa presenta valori alti di questa voce di bilancio, ciò potrebbe comportare una costante presenza di costi fissi che incidono sui costi di produzione, variabile da cui *Ebitda Margin* è formato, e viceversa.

Le disponibilità liquide comprendono depositi bancari e postali, assegni, denaro e valori in cassa esistenti in cassa alla fine dell'esercizio: in questa voce di bilancio rientrano quindi, quegli elementi patrimoniali che sono immediatamente disponibili ed utilizzabili. Insieme a crediti e rimanenze, le disponibilità liquide costituiscono il capitale circolante, che evidenzia la capacità più o meno sostanziosa di far fronte agli impegni di breve periodo, rappresentati da debiti di funzionamento e di finanziamento. La gestione della liquidità e il margine operativo lordo (*Ebitda*), seppur non direttamente correlati, rappresentano due aspetti chiave da monitorare congiuntamente per il buon andamento dell'attività imprenditoriale.

Conclusioni

Lo studio effettuato per prevedere valori dell'*Ebitda Margin* di imprese manifatturiere del settore alimentare italiano, attraverso l'utilizzo di dati storici, non mostra buoni risultati. È consigliabile, quindi, ragionare su modelli più complessi che includano un maggior numero di attributi capaci di descrivere il contesto in cui l'impresa opera e fattori che incidono sull'andamento della stessa, soprattutto nel caso delle grandi imprese.

Oltre a ciò, una clusterizzazione più ridotta potrebbe favorire l'ottenimento di risultati migliori.

È necessario, però, sempre considerare che il range della variabile è molto ristretto e ciò rende difficile un'accurata predizione dei valori delle istanze.

Inoltre, la stima dei valori dell'indice di redditività è fortemente influenzato da variabili che dipendono in modo significativo da decisioni e dinamiche sociali.

Motivi per cui, sebbene modelli più articolati potrebbero produrre *perfor-*

mances migliori, la predizione dei valori dell'*Ebitda Margin* risulta essere un obiettivo complesso da perseguire.

Bibliografia

- [1] Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar (2020)

Introduction to Data Mining, Pearson.

- [2] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2020)

Introduzione all'apprendimento statistico, Piccin.