

Università degli studi di Milano-Bicocca

Big data in behavioural psychology

Progetto Finale

Ricerca dei comportamenti nelle offerte di vendita del mercato elettrico

Analisi delle curve di offerta nel MGP

Autori:

Beatrice Fumagalli - 784549 - b.fumagalli9@campus.unimib.it

Matteo Porcino - 748876 - m.porcino1@campus.unimib.it

Nicolò Monti – 769709 – n.monti1@campus.unimib.it



Introduzione

La psicologia ha a lungo studiato la personalità e il comportamento dell'uomo, al fine di estrarre pattern comportamentali e capire quali siano le ragioni alla base delle sue azioni. Solo recentemente, grazie all'avvento dell'era dei Big Data, è stato possibile applicare questi studi anche in altri ambiti. L'analisi del comportamento diventa così un importante strumento utilizzato nel campo dell'industria energetica, più precisamente all'interno del *mercato elettrico* italiano (d'ora in avanti indicato con *ME*), oggetto del caso di studio considerato, al fine di indagare quelle che sono le strategie che i principali operatori mettono in atto all'interno di esso. Grazie a questo strumento è possibile individuare i diversi *modus operandi* dei player e i loro rispettivi profili di avversione e propensione al rischio.

Nella seguente relazione sono stati considerati due articoli nei quali vengono trattate le metodologie utilizzate per l'estrazione di questi profili:

- *A Cluster Analysis of PowerProducer's Bidding Using Average Electricity Price Difference-Integration Model*¹.
- *Analisi delle curve di offerta della borsa del mercato elettrico*²

Workflow

Di seguito viene presentato sinteticamente il flusso di lavoro che è stato seguito al fine di estrarre le strategie comportamentali degli operatori del *ME*.

- 1) Esplorazione iniziale del dataset per poter individuare eventuali anomalie all'interno dei dati
- 2) Pre-processing e messa in qualità del dataset fornito
- 3) Arricchimento del dataset con dati necessari all'analisi delle strategie
- 4) Creazione delle curve aggregate d'offerta
- 5) Rappresentazione vettoriale delle curve d'offerta
- 6) Clusterizzazione delle curve di offerta
- 7) Interpretazione dei risultati e individuazione dei pattern comportamentali

Il Dataset

Si è scelto di utilizzare i dati forniti dalla società Cura Consorzio all'interno del file "DATA_2.csv". Il dataset è composto da 2.552.529 osservazioni e da 18 attributi:

- **DATE:** timestamp dell'osservazione;
- **INTERVAL NO:** indica l'orario a cui l'offerta si riferisce, i valori variano da 1 a 24;
- **BID_OFFER_DATE:** indica il giorno dell'offerta;
- **PURPOSE CD:** identifica la tipologia di offerta;
- **TRANSACTION REFERENCE NO:** numero della transazione (univoco);

¹ Haol, Z. Jianhua, C. Zhenxiang, H Dongming, W. Weizhen, S. A Cluster Analysis of PowerProducer's Bidding Using Average Electricity Price Difference-Integration Model, IEEE International Conference on Electricity Utility Deregulation, Restructuring and Power Technologies (DRPT2004), Hong Kong, April 2004.

² L. Passamonti, S. Caruggi, R. Turra, G. Pedrazzi, Analisi delle curve di offerta della borsa del mercato elettrico, https://www.cineca.it/sites/default/files/borsa_el62.pdf

- **STATUS_CD**: indica lo stato dell'offerta dopo che si è conclusa la sessione di mercato;
- **MARKET_CD**: indica il mercato nel quale viene fatta l'offerta;
- **SCOPE**: presente solo se l'attributo **MARKET_CD** assume valore *msd*;
- **ZONE_CD**: indica la zona di offerta in acquisto/vendita;
- **OPERATORE**: società titolare dell'impianto o incaricata del dispacciamento;
- **UNIT_REFERENCE_NO**: codice dell'unità di produzione/consumo;
- **QUANTITY_NO**: quantità di energia (MWh) presentata dall'operatore;
- **AWARDED QUANTITY_NO**: quantità di energia (MWh) riconosciuta nel mercato;
- **ENERGY PRICE_NO**: prezzo (€/MWh) unitario presentato dall'operatore;
- **AWARDED PRICE_NO**: prezzo (€/MWh) unitario riconosciuto nel mercato;
- **MERIT ORDER_NO**: ordine di merito dell'offerta.

Pre-processing

Prima di poter effettuare l'analisi vera e propria dei dati è stato necessario applicare delle operazioni di *pre-processing* al dataset iniziale. Vengono riportate sotto le azioni che sono state eseguite per la messa in qualità dei dati.

- Correzione di alcune osservazioni, circa una ventina, al cui interno sono presenti errori a livello di formattazione dei valori degli attributi. L'operazione è stata effettuata correggendo manualmente le osservazioni anomale poiché in numero esiguo;
- Eliminazione di alcune osservazioni con attributi "PURPOSE_CD", "STATUS_CD", "MARKET_CD" e "OPERATORE" con valori non riconducibili a nessuno di quelli corretti.
- All'interno di un numero elevato di attributi sono presenti errori sintattici. Per esempio, è presente all'interno dell'attributo "OPERATORE", che rappresenta la società che opera nel ME, il valore "ENEL PRODUZIONE S.P.A." scritto in una decina di modi differenti. È stato quindi necessario applicare delle funzioni (definite ad-hoc per questo task) per eliminare questo tipo di ambiguità nei dati.

Attributo	Modalità prima del pre-processing	Modalità dopo il pre-processing
OPERATORE	36	11
TYPE_CD	10	2
ZONE_CD	13	1
UNIT_REF_NO	53	30
SCOPE	12	6

- È stato necessario cambiare il formato di alcuni attributi in quanto interpretati dall'ambiente *Python* come "*str*" (string) quando invece si riferiscono a dati di tipo quantitativo, per cui "*float*".

- Sono state eliminate quelle colonne contenenti informazioni ridondanti o superflue ai fini dell'analisi:
 - BID_OFFER_DATE_DT, ridondante con l'attributo Date
 - TRANSACTION_REFERENCE_NO, non necessario
- Rimozione delle osservazioni con valore dell'attributo "OPERATORE" pari a "Bilateralista" in quanto non rappresenta un operatore del ME.

Al termine della fase di *pre-processing* si ottiene così un nuovo dataset di qualità superiore a quello di partenza con un totale di 2.450.249 osservazioni per le successive analisi.

La fase di *pre-processing* è stata implementata servendosi principalmente delle funzioni messe a disposizione dalla libreria *Pandas*, utilizzando come linguaggio di programmazione *Python* 3.6.

Focus dell'Analisi

Come accennato all'interno dell'introduzione è possibile effettuare un elevato numero di analisi e studi su questi dati. Al fine di ottenere dei risultati il più verosimili ed utili possibili si è deciso di analizzare non la totalità del ME, ma solamente quello che viene chiamato *Mercato del Giorno Prima* (a cui ci riferiremo con MGP d'ora in avanti). All'interno di questa specifica sezione del ME, gli operatori piazzano le loro offerte di vendita e acquisto di energia elettrica; questi volumi di energia sono quindi venduti mediante un sistema ad aste.

Filtrando i dati puliti e considerando solamente le osservazioni riguardanti le offerte di vendita del MGP si ottengono un totale di 300.528 record.

Creazione curve aggregate di offerta

L'analisi delle strategie di offerta degli operatori non può essere fatta analizzando semplicemente i dati che sono stati forniti. Secondo la metodologia espressa all'interno degli articoli ricercati ^[1,2] è stato infatti necessario calcolare le *curve di offerta* di questi operatori.

Per poterle calcolare è stato necessario ordinare i dati, per ogni operatore e data, in ordine crescente di prezzo ed aggiungere dei nuovi attributi all'interno del dataset, rispettivamente:

- QUANTITY_CUMSUM: rappresenta la quantità cumulata dell'energia elettrica offerta dall'operatore;
- QUANTITY_NORM: valore normalizzato della quantità cumulata per poter rendere confrontabili le diverse offerte dei diversi operatori.

Gli articoli considerati suggerivano anche la normalizzazione del prezzo delle offerte, in quanto precedentemente al 2016 era presente un prezzo limite a cui gli operatori potevano offrire la propria energia, questa limitazione non è più presente all'interno del ME per cui si è optato per mantenere i prezzi originali al fine di ottenere una leggibilità maggiore dei risultati finali. Una volta creati questi nuovi attributi è stato possibile mostrare i grafici a gradini delle diverse curve di offerta per i diversi operatori nei diversi giorni e orari. Si ottengono così un totale di 87.260

curve di offerta. È stata inserita una retta di riferimento sull'asse delle y indicante il prezzo di riferimento medio annuo dell'energia elettrica.

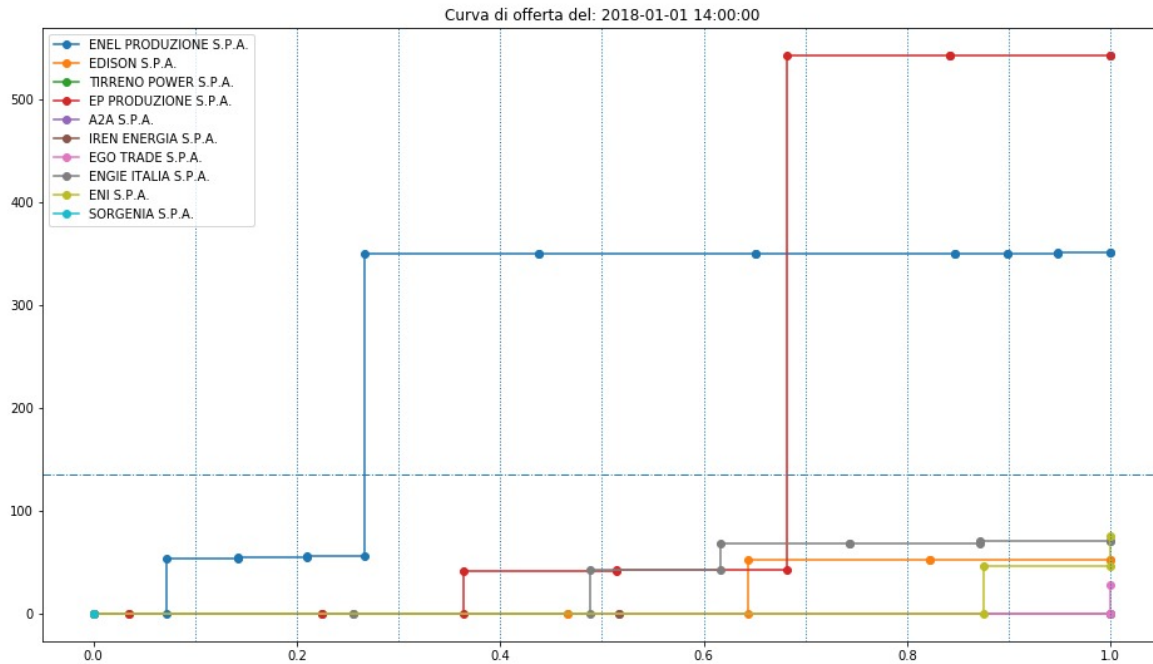


Figura 1 - rappresentazione di una curva d'offerta con i diversi operatori del ME

Rappresentazione vettoriale delle curve

Le curve d'offerta appena calcolate sono rappresentate mediante coppie di punti dove la prima dimensione rappresenta la quantità cumulata di energia offerta mentre la seconda rappresenta il prezzo alla quale viene offerta la suddetta quantità. Individuate in questo modo, le curve non sono confrontabili tra loro, in quanto ognuna di esse è formata da un numero variabile di punti, equivalenti al numero di offerte fatte dall'operatore.

Per poter applicare le tecniche di *Cluster Analysis* è stato necessario rappresentare in maniera univoca tutte le curve d'offerta calcolate, permettendone così il confronto tra loro.

Ogni curva viene segmentata in un numero di intervalli pari a 10 e per ogni intervallo viene calcolata l'area compresa tra la funzione a gradini della curva d'offerta e la retta di riferimento del prezzo medio annuo.

$$s_{i,j} = \int_{Q^{*}_{j-1}}^{Q^{*}_j} (P - \bar{P}) dQ * \quad j \in [1,10] \quad (1)$$

$$S_i = [s_{i,1}, s_{i,2}, \dots, s_{i,10}] \quad (2)$$

La formula 1 rappresenta l'area di un singolo segmento di curva, compreso tra il segmento $j-1$ e j , dato un operatore tra quelli presenti e i , l'intervallo di tempo considerato.

Nella 2 è possibile invece vedere come ogni curva al tempo i , indicata con S_i , viene così rappresentata come l'insieme delle aree per ognuno dei 10 segmenti.

Per poter calcolare le quantità s_{ij} per ogni operatore è stato necessario implementare una funzione ad-hoc che fosse in grado di calcolare l'integrale sopra descritto.

Alla fine di questa fase ogni curva è rappresentata all'interno dello stesso spazio vettoriale definito in \mathbb{R}^{10} ed è pronta per essere inserita all'interno degli algoritmi di clusterizzazione scelti.

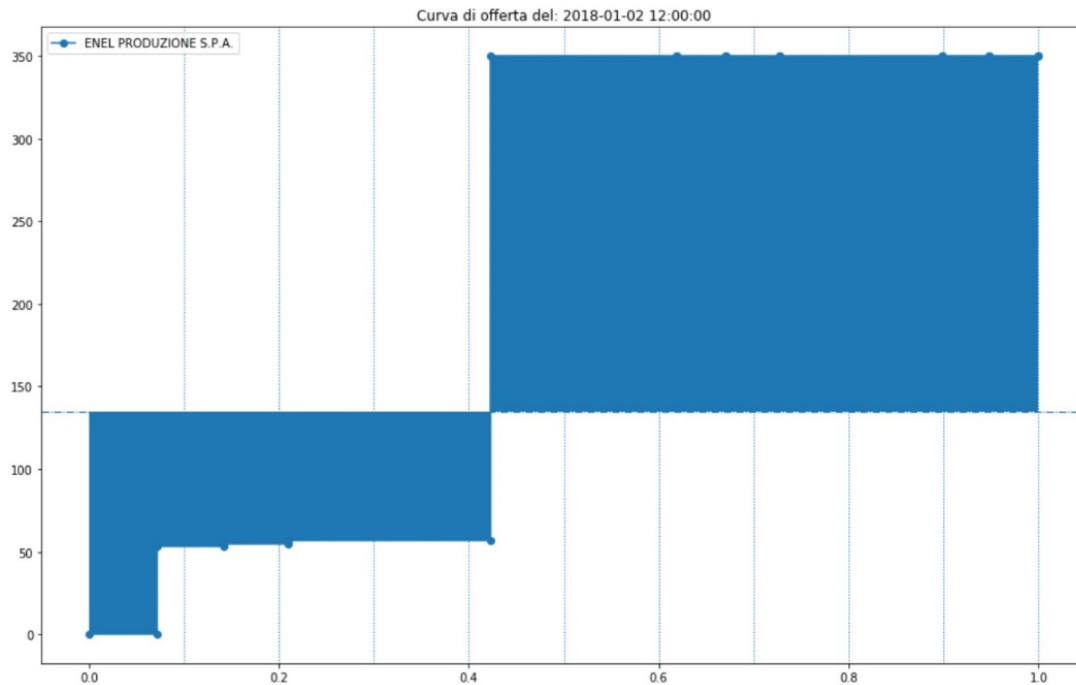


Figura 2- curva di offerta per l'operatore ENEL con evidenziata l'area calcolata dalla funzione

Clustering

L'algoritmo scelto per la clusterizzazione delle strategie d'offerta è quello conosciuto come **K-Means**, un algoritmo di clustering agglomerativo in grado di, a seconda della distanza utilizzata, individuare gruppi di osservazioni "simili" tra loro e raggrupparle in *cluster*. I cluster risultanti potranno essere poi considerati come le diverse strategie di comportamento degli operatori energetici.

Prima di effettuare il clustering vero e proprio è necessario individuare il numero più consono di cluster in cui dividere le osservazioni, in quanto il *k-means* è un algoritmo di clustering *top-down* che non è in grado di stabilire in maniera autonoma il numero ottimale di cluster.

Per la scelta di questo numero sono stati utilizzati e confrontati tre differenti metodi:

- L'indice di *Silhouette*
- Il metodo dell'*Elbow*
- Il decremento della *varianza intra-cluster*

Una volta individuato il numero di cluster da passare all'algoritmo, il *k-means* è stato applicato inizialmente con il metodo *Ward* e successivamente con il metodo *Average Linkage*. I risultati

ottenuti hanno mostrato delle performance leggermente migliori con il metodo *Ward*, ottenendo così il seguente numero di *clusters* per ciascun operatore:

Operatore nel <i>ME</i>	Numero di cluster individuati
ENEL PRODUZIONE S.P.A.	8
EDISON S.P.A.	4
TIRRENO POWER S.P.A.	4
EP PRODUZIONE S.P.A.	4
A2A S.P.A.	4
IREN ENERGIA S.P.A.	4
EGO TRADE S.P.A.	3
ENGIE ITALIA S.P.A.	4
ENI S.P.A.	3
SORGENIA S.P.A.	4

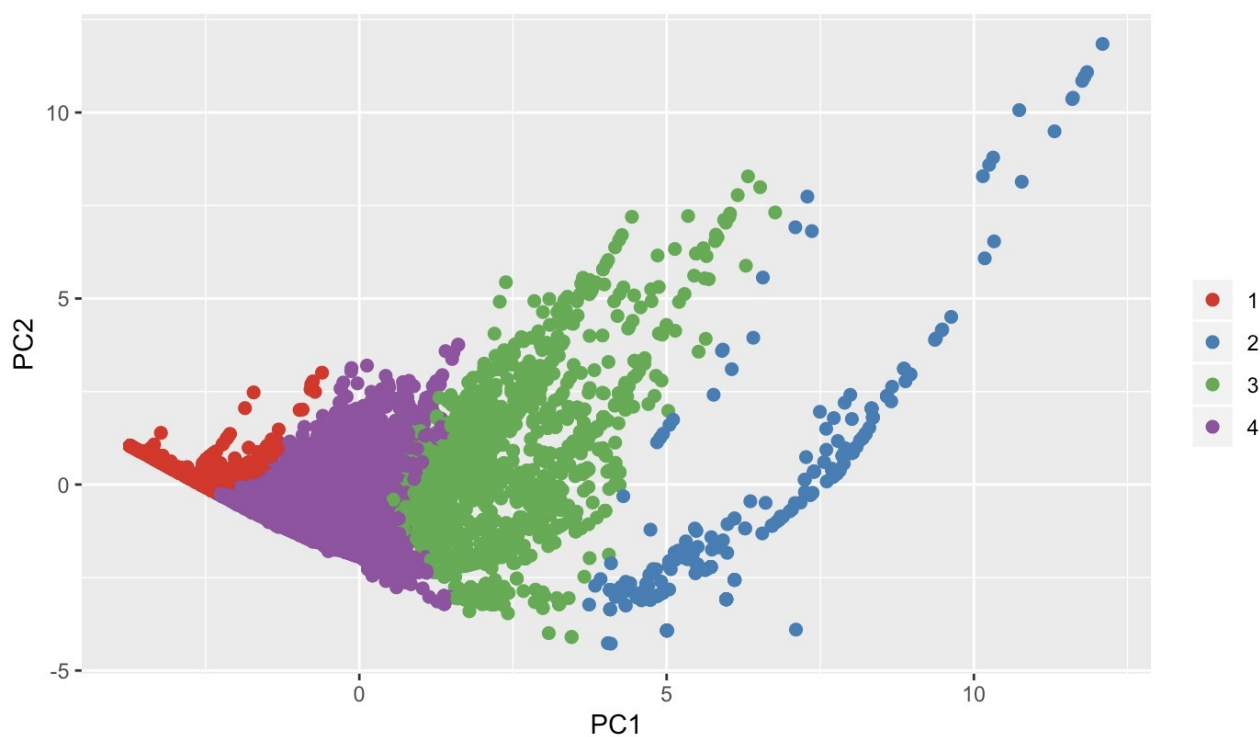


Figura 3- Visualizzazione dei clusters K-Means dell'operatore A2A con le prime due componenti principali

I *cluster* ottenuti definiscono delle tipologie di comportamento, interpretabili in termini di strategie di offerta di vendita e costituiscono la base per la successiva creazione di un modello di classificazione.

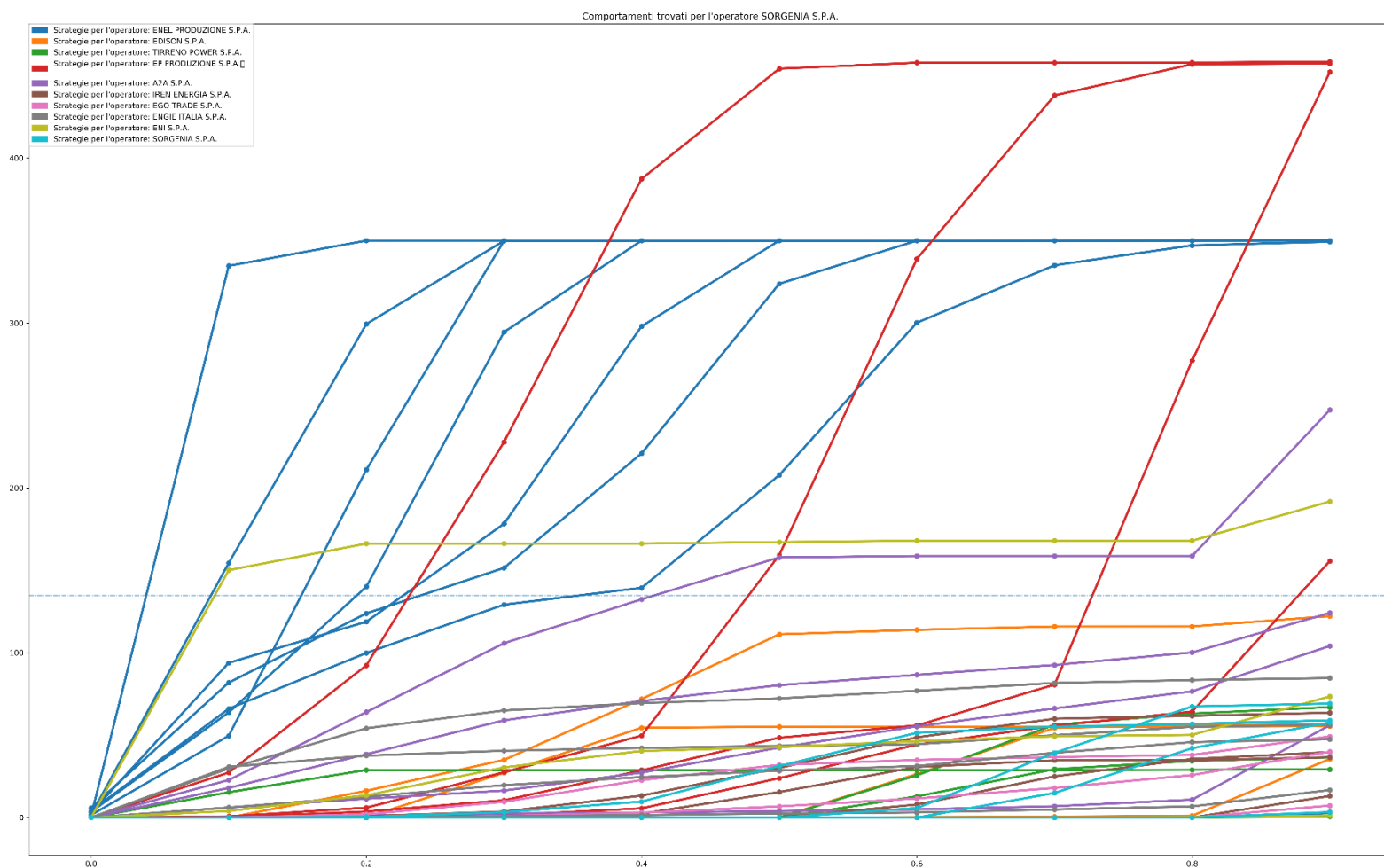


Figura 4- rappresentazione grafica dei cluster individuati per ogni operatore

Risultati e Conclusioni

L' MGP è un mercato con bassa propensione al rischio, ma analizzando le curve dei centroidi emergono diversi comportamenti tra gli operatori che possono essere così sintetizzati:

- Enel mostra in gran parte delle curve ad S e qualche curva piatta al di sopra del prezzo medio annuo di riferimento. In particolare, tende ad offrire una grande quantità di energia ad un unico livello di prezzo, di gran lunga superiore alla media. Pertanto, tra gli operatori è uno dei più propensi al rischio, assieme a EP Produzione. Questo comportamento può essere dovuto al fatto che, essendo *MGP* un mercato a bassa propensione al rischio, ENEL può far valere il suo ruolo di principale operatore nel mercato Elettrico in modo da influenzare le variazioni del prezzo equilibrio, contribuendo ad aumentare il costo dell'energia venduta durante le sedute di *MGP*.
- A2A ed Eni presentano una retta inclinata e una piatta propense al rischio, mentre le altre curve tendono ad essere a prezzo zero.

- Tutti gli operatori restanti presentano una tendenza a offerte a prezzo zero. Pertanto, questi operatori, lasciando che sia il mercato a determinare il prezzo di vendita, risultano essere i più avversi al rischio e contribuiscono a far diminuire il prezzo di equilibrio di vendita.

Sviluppi futuri

L'approccio utilizzato per analizzare le strategie degli operatori all'interno del *MGP* può essere applicato anche alle altre sessioni di mercato presenti nel *ME*, ovvero il *Mercato Infra-giornaliero*, indicato con *MI*, e il *Mercato del Servizio di Dispacciamento*, abbreviato *MSD*.

La stessa analisi può essere effettuata non considerando i singoli operatori del *ME*, ma considerando i diversi impianti di produzione potendo così individuare se esistono diverse strategie in base alle diverse caratteristiche degli impianti.