

# CarlCon:

## Tempo di accelerazione di un'auto

### Componenti del gruppo

- Giuseppe Caggiano [MAT. 735400], g.caggiano8@studenti.uniba.it

Link progetto GitHub:

[CarlCon](#)

Anno Accademico 2023-2024

## INDICE

- *Capitolo 0 : Introduzione*
- *Capitolo 1 : Dataset e Features*
- *Capitolo 2 : Clustering (Apprendimento non supervisionato)*
- *Capitolo 3 : Knowledge Base attraverso Prolog*
- *Capitolo 4 : Regressione sull'accelerazione*
- *Capitolo 5 : KNN*
- *Capitolo 6 : AdaBoost*
- *Capitolo 7 : Random Forest*
- *Capitolo 8 : Alberi di Decisione*
- *Capitolo 9 : Risultati finali*
- *Capitolo 10: Conclusione*

## Capitolo 0: Introduzione

L'obiettivo di questo progetto è stato analizzare ed effettuare operazioni sul tempo di accelerazione da 0 a 100 km/h di un'auto tramite un dataset di Kaggle chiamato `Car_specification_dataset_1945_2020`. Questo dataset contiene modelli di auto con varie caratteristiche per singola auto, tra cui il target di regressione scelto, ovvero il tempo di accelerazione da 0 a 100 km/h. Tra le variabili rilevanti per l'analisi dell'accelerazione vi sono, ad esempio, la lunghezza, la larghezza, la massa, la potenza del motore e la coppia massima, che influenzano direttamente le prestazioni di accelerazione.

Ho utilizzato tecniche di :

- **apprendimento non supervisionato** per l'implementazione del k-means per individuare i cluster di appartenenza di ogni occorrenza del dataset (creando una nuova caratteristica con il cluster ottenuto per migliorare accuratezza)

- **apprendimento supervisionato** con regressione tramite KNN e Alberi di Decisione

Inoltre, il progetto utilizza **Prolog** per ottenere nuove informazioni attraverso il ragionamento automatico.

## Requisiti funzionali

Il progetto è stato realizzato con Python versione 3.12 per la grande vastità di librerie che aiutano nell'implementazione. L'IDE che ho usato per la stesura del codice è stato PyCharm, semplice e funzionale. Le librerie che ho utilizzato sono le seguenti:

- **pandas**: per la manipolazione e analisi dei dati.
- **scikit-learn**: per il preprocessing, clustering e apprendimento supervisionato.
- **matplotlib**: per la visualizzazione dei dati attraverso i grafici
- **pyswip**: per l'integrazione con Prolog.

## Configurazioni iniziali

Installare SWI-Prolog dal sito ufficiale <https://www.swi-prolog.org/download/stable>, fondamentale per l'utilizzo della libreria pyswip. Successivamente, dopo aver aperto il progetto con l'IDE preferito, avviare lo script `install.py` per scaricare tutte le librerie importanti per il corretto funzionamento del progetto.

## Capitolo 1 : Dataset e Features

Il dataset è stato trovato sul sito di Kaggle, una sorta di archivio di dataset, e contiene modelli di auto, con annesse caratteristiche, prodotte dal '45 al 2020. Molte colonne contengono dati mancanti o

presentano un alto grado di sparsità, aspetto che ha richiesto un significativo lavoro di preprocessing per ridurre la dimensione del dataset ed eliminare valori mancanti o irrilevanti, tra cui:

- maximum\_torque\_n\_m (coppia massima del motore)
- injection\_type (tipo di iniezione del motore)
- engine\_hp (numero indicante i cavalli del motore)
- boost\_type (tipo di turbina)
- transmission (tipo di trasmissione manuale, automatica o sequenziale)

oltre ovviamente alle dimensioni del mezzo e ad altre caratteristiche.

I dati sono poi preprocessati per renderli pronti all'analisi.

Per prima cosa, come detto precedentemente, ho deciso di 'scremare' le caratteristiche non importanti per l'obiettivo alleggerendo il mio dataset. Così ho eliminato le colonne che non avevano impatto sul tempo di accelerazione, oppure quelle con troppi valori vuoti.

Successivamente, oltre a questo sono stati fatti alcuni aggiustamenti nelle colonne che erano fondamentali per il tempo di accelerazione, al fine di rendere più snello e leggibile il dataset, ma anche per eliminare dei problemi sintattici che davano problemi nella base di conoscenza Prolog.

Dopo queste fasi viene creato un nuovo file chiamato preprocessed.csv

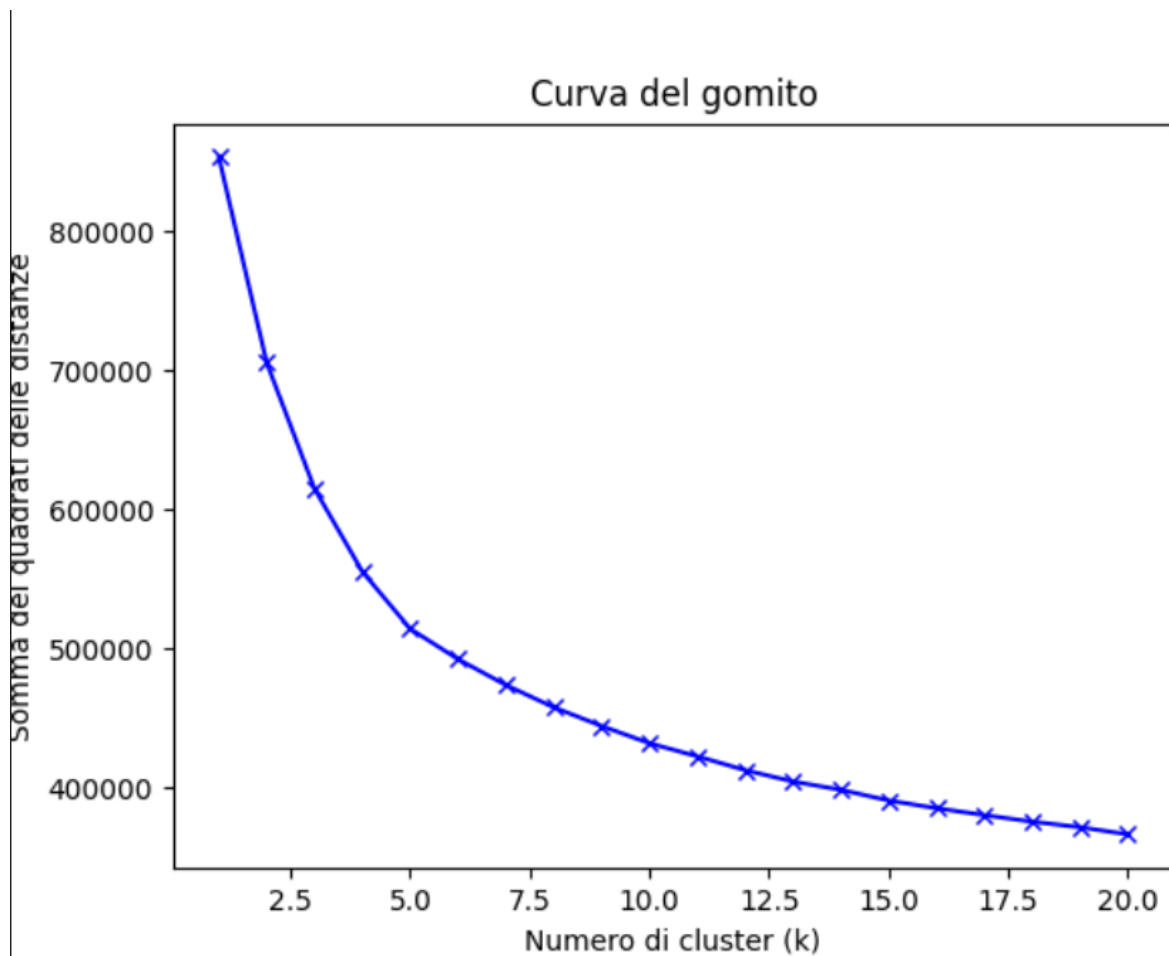
## Capitolo 2 : Clustering (Apprendimento non supervisionato)

L'apprendimento non supervisionato è una branca dell'apprendimento automatico in cui l'agente viene addestrato su un insieme di dati senza etichette.

- Clustering : tecnica di machine learning che suddivide i dati in gruppi omogenei chiamati **cluster**. L'obiettivo di questa fase è ottenere un secondo dataset che aggiunge a preprocessed.csv una nuova feature con il cluster ottenuto.

Personalmente ho applicato l'algoritmo K-Means sfruttando la distanza euclidea. Inoltre ho deciso di escludere la feature ID perché può non essere un dato importante allo scopo del mio progetto.

Ora andiamo con l'individuazione del numero ottimale di cluster utilizzabili. Per fare ciò ho implementato la tecnica della curva del gomito con un intervallo da 1 a 20.



**Numero ottimale di cluster: 6**

Quindi ottenuto il numero ottimale di cluster questo viene usato per calcolare il cluster di appartenenza per ogni occorrenza, poi inserito nella caratteristica a cui appartiene.

Non siamo sicuri però che sapere il cluster di appartenenza possa aiutare nella regressione quindi otteniamo un nuovo file chiamato clustered.csv

## Capitolo 3 : Knowledge Base attraverso Prolog

A questo punto creo una Knowledge Base sfruttando il Prolog con la libreria Pyswip di Python, per ottenere nuova conoscenza utile per creare due nuovi dataset per la valutazione dell'apprendimento automatico sfruttando ragionamento automatico e la struttura di individui e relazioni:

- **Individui:** Knowledge Base popolata con fatti del dataset preprocessato modellando il dominio, e trova la classe di individui con simbolo di funzione annesso car(C) cioè l'auto con ID C.

- **Proprietà:** le proprietà associate a individuo car(C) corrispondono alle feature nel dataset che sono usate ai fini dell'ottenere nuove informazione nel ragionamento automatico

- engine\_hp(car(C), H)

- engine\_hp(car(C), H)

- curb\_weight\_kg(car(C), K)
- height\_mm(car(C), HeightMm)
- width\_mm(car(C), WidthMm)
- capacity\_cm3(car(C), Cap)
- number\_of\_cylinders(car(C), N)
- max\_speed\_km\_per\_h(car(C), M)
- make(car(C), Manufacturer)
- cylinder\_layout(car(C), Cylinder\_layout)
- max\_trunk\_capacity\_l(car(C), Capacity)
- model(car(C), MaxModel)
- year\_from(car(C), Year)

- **Clausole Definite:** clausole definite per effettuare ragionamento . Ad esempio:

- **Clausole rapporti:**

- calcolo coefficiente aerodinamico tramite caratteristiche dell'auto e densità dell'aria
- calcolo potenza specifica con rapporto cavalli/cilindrata
- calcolo rapporto lunghezza/altezza
- calcolo rapporto lunghezza/potenza

- **Clausole di Categorizzazione:**

- classificazione potenza auto in bassa, media e alta
- classificazione tipo auto in sportiva, coupè , berlina e suv
- calcolo capacità media del bagagliaio
- calcolo modello più potente di un marchio

- **Clausole ausiliarie :** clausole per capire quale sia l'elemento più frequente di una lista o la forma dell'auto in base alle caratteristiche 'fisiche' dell'auto.

- **Query :** dopo le clausole, si effettuano query per il calcolo dei valori da inserire nelle occorrenze delle nuove feature, in cui va sostituita la variabile C in car(C) in ogni clausola sfruttata con ID auto. Con esse creo due nuovi dataset **prolog\_generated.csv** ( con le feature create in questa fase) e **clustered\_plus\_prolog\_generated.csv** (con feature di clustered.csv + le feature create in questa fase)

## Capitolo 4 : Regressione sull'accelerazione

Dopo aver generato i vari dataset, procediamo col target su tempo di accelerazione.

Usiamo la libreria scikit-learn di Python per svolgere il training tramite i regressori KNN, AdaBoost, Random Forest e Alberi di decisione per ogni dataset generato.

In ogni fase di addestramento, si è cominciato con l'encoding delle variabili categoriche (differenti in base al dataset) utilizzando la classe **OrdinalEncoder**.

Successivamente, sono stati definiti gli iperparametri da ottimizzare per ciascun tipo di regressore, insieme ai rispettivi valori e intervalli da testare.

L'addestramento dei modelli è stato condotto con un approccio di **10-fold cross-validation**, applicato per ogni combinazione di iperparametri, e le metriche di valutazione riportate rappresentano la media dei risultati ottenuti sui 10 modelli addestrati per ciascuna combinazione. Per garantire un comportamento deterministico, dove possibile (nei modelli AdaBoost, Decision Tree e Random Forest), il parametro *random\_state* è stato fissato a 0.

Le metriche di valutazione selezionate sono le seguenti:

- *neg\_mean\_absolute\_error*
- *neg\_mean\_squared\_error*
- *max\_error*

Infine, per ottenere una comprensione più approfondita dei fattori che influenzano la variabile obiettivo calcola la **feature importance** sul modello migliore per ogni regressore e per ogni dataset usato.

## Capitolo 5 : KNN

Il primo modello di regressione analizzato sui quattro dataset precedentemente generati è il K-Nearest Neighbors (KNN). Le prestazioni del KNN sono state valutate utilizzando la *10-fold Cross-Validation*, testando tutte le combinazioni dei seguenti iperparametri:

- **n\_neighbors**: sono stati testati i valori 3, 5 e 7, rappresentanti il valore di kkk.
- **weights**: sono stati valutati i metodi *uniform* e *distance* per il calcolo delle distanze. Con *uniform*, tutti i vicini hanno lo stesso peso, mentre *distance* assegna un peso maggiore ai vicini più prossimi.
- **algorithm**: è stato testato *kd-tree* come tipo di algoritmo per la ricerca dei kkk punti più vicini.
- **p**: è stato utilizzato il valore 2, specificando l'uso della distanza euclidea per identificare i vicini.
- **leaf\_size**: sono stati provati i valori 20, 30 e 40, i quali definiscono la dimensione massima delle foglie dell'albero kd-tree.

Questa configurazione ha permesso di ottimizzare il modello in modo completo, esplorando le diverse combinazioni di parametri per individuare la configurazione migliore.

## Valutazioni dei dataset

-preprocessed.csv

```
Migliori parametri: {'algorithm': 'kd_tree', 'leaf_size': 20, 'n_neighbors': 7, 'p': 2, 'weights': 'distance'}
Migliore score: -4.497272107556813
Risultati della ricerca a griglia:
mean_test_neg_mean_absolute_error: [-1.56477676 -1.56086766 -1.52441846 -1.51809623 -1.50571768 -1.49723788
-1.56477676 -1.56086766 -1.52441846 -1.51809623 -1.50571768 -1.49723788
-1.56478285 -1.56087332 -1.5244212 -1.518099 -1.50571964 -1.49723912]
mean_test_neg_mean_squared_error: [-4.84635823 -4.82986309 -4.62930042 -4.60052566 -4.53561283 -4.49727211
-4.84635823 -4.82986309 -4.62930042 -4.60052566 -4.53561283 -4.49727211
-4.84635905 -4.82986441 -4.62930884 -4.60053409 -4.53561754 -4.49727397]
mean_test_max_error: [-14.24666667 -14.2206763 -13.852 -13.84469045 -14.06428571
-14.03863873 -14.24666667 -14.2206763 -13.852 -13.84469045
-14.06428571 -14.03863873 -14.24666667 -14.2206763 -13.852
-13.84469045 -14.06428571 -14.03863873]
```

Feature importance:

```
Feature importance in ordine decrescente:
max_speed_km_per_h: 0.3741932500103024
engine_hp: 0.16522704958788387
maximum_torque_n_m: 0.09137648308043811
number_of_gears: 0.03611993179591277
rear_track_mm: 0.03345764508316736
number_of_cylinders: 0.027630012841870204
front_track_mm: 0.026811846397810448
capacity_cm3: 0.024573592127084192
curb_weight_kg: 0.02438423132235327
width_mm: 0.022597342651456666
valves_per_cylinder: 0.02164166510129761
cylinder_layout: 0.0174863946086766
length_mm: 0.01559989076755003
wheelbase_mm: 0.014648319699286507
turnover_of_maximum_torque_rpm: 0.012818707812711636
cylinder_bore_mm: 0.012628068913964027
fuel_tank_capacity_l: 0.011594622842667907
year_from: 0.010830589004631226
```

-clustered\_plus\_prolog\_generated.csv

```
Migliori parametri: {'algorithm': 'kd_tree', 'leaf_size': 20, 'n_neighbors': 7, 'p': 2, 'weights': 'distance'}
Migliore score: -4.573551172300965
Risultati della ricerca a griglia:
mean_test_neg_mean_absolute_error: [-1.57509624 -1.57120925 -1.53694865 -1.53079687 -1.51781609 -1.50983773
-1.57509624 -1.57120925 -1.53694865 -1.53079687 -1.51781609 -1.50983773
-1.57510233 -1.5712149 -1.53694043 -1.53078925 -1.51782523 -1.50984543]
mean_test_neg_mean_squared_error: [-4.96070225 -4.94342168 -4.74195234 -4.71158899 -4.60902202 -4.57355117
-4.96070225 -4.94342168 -4.74195234 -4.71158899 -4.60902202 -4.57355117
-4.9607059 -4.94342555 -4.7419452 -4.71158243 -4.60903544 -4.57356107]
mean_test_max_error: [-15.45666667 -15.4613417 -14.894 -14.89490388 -14.41142857
-14.41096821 -15.45666667 -15.4613417 -14.894 -14.89490388
-14.41142857 -14.41096821 -15.45666667 -15.4613417 -14.894
-14.89490388 -14.41142857 -14.41096821]
```



## Feature importance:

```
Feature importance in ordine decrescente:  
weight_to_power_ratio: 0.25851917347830666  
max_speed_km_per_h: 0.16775473286833928  
length_power_ratio: 0.16717654537545876  
engine_hp: 0.07407300790561075  
max_speed_power_ratio: 0.06380548937684422  
maximum_torque_n_m: 0.04096502945786751  
specific_power: 0.0396811140850274  
number_of_gears: 0.016192941774014244  
rear_track_mm: 0.014999410901121056  
number_of_cylinders: 0.01238682264661136  
front_track_mm: 0.01202003010489313  
capacity_cm3: 0.011016597393942933  
curb_weight_kg: 0.010931704972146187  
width_mm: 0.010130624162581961  
valves_per_cylinder: 0.009702183959209007  
length_ratio: 0.0096706991547086  
number_cylinder_capacity_ratio: 0.008366985428300262  
cylinder_layout: 0.007839332901724302
```

## -prolog\_generated.csv

```
Migliori parametri: {'algorithm': 'kd_tree', 'leaf_size': 40, 'n_neighbors': 3, 'p': 2, 'weights': 'uniform'}  
Migliore score: -3.5877243865793615  
Risultati della ricerca a griglia:  
mean_test_neg_mean_absolute_error: [-1.25478362 -1.26101679 -1.26223984 -1.25978518 -1.27119598 -1.26412282  
-1.25478362 -1.26101679 -1.26223984 -1.25978518 -1.27119598 -1.26412282  
-1.25468768 -1.26086292 -1.26242439 -1.25990258 -1.27124688 -1.26418851]  
mean_test_neg_mean_squared_error: [-3.5887855 -3.60154995 -3.70507413 -3.66465592 -3.90731574 -3.83601527  
-3.5887855 -3.60154995 -3.70507413 -3.66465592 -3.90731574 -3.83601527  
-3.58772439 -3.60020188 -3.7044486 -3.66397626 -3.90725859 -3.83597841]  
mean_test_max_error: [-12.46 -12.4629665 -12.568 -12.542415 -12.84571429  
-12.80292201 -12.46 -12.4629665 -12.568 -12.542415  
-12.84571429 -12.80292201 -12.46 -12.4629665 -12.568  
-12.542415 -12.84571429 -12.80292201]
```

## Feature importance:

```
Migliori parametri: {'algorithm': 'kd_tree', 'leaf_size': 20, 'n_neighbors': 7, 'p': 2, 'weights': 'distance'}  
Migliore score: -4.497168784094539  
Risultati della ricerca a griglia:  
mean_test_neg_mean_absolute_error: [-1.56478133 -1.56087195 -1.52442394 -1.51810138 -1.50569288 -1.49721355  
-1.56478133 -1.56087195 -1.52442394 -1.51810138 -1.50569288 -1.49721355  
-1.56478742 -1.56087762 -1.52442668 -1.51810415 -1.50569484 -1.49721479]  
mean_test_neg_mean_squared_error: [-4.8463657 -4.82986972 -4.62932761 -4.60055114 -4.53550867 -4.49716878  
-4.8463657 -4.82986972 -4.62932761 -4.60055114 -4.53550867 -4.49716878  
-4.84636651 -4.82987103 -4.62933603 -4.60055957 -4.53551338 -4.49717064]  
mean_test_max_error: [-14.24666667 -14.22067589 -13.852 -13.84468995 -14.06428571  
-14.03863935 -14.24666667 -14.22067589 -13.852 -13.84468995  
-14.06428571 -14.03863935 -14.24666667 -14.22067589 -13.852  
-13.84468995 -14.06428571 -14.03863935]
```

-clustered.csv

```
Feature importance in ordine decrescente:  
weight_to_power_ratio: 0.46880846942532944  
length_power_ratio: 0.3031642849030678  
max_speed_power_ratio: 0.11570729324725515  
specific_power: 0.0719592365587329  
length_ratio: 0.017537212455045262  
number_cylinder_capacity_ratio: 0.015173008560909055  
avg_trunk_capacity_brand: 0.0053020630873005575  
shape: 0.0014408536139783607  
most_powerful_car_brand: 0.0003252803480765833  
drag_coefficient: 0.00021651319990577023  
power: 0.0002067472976678387  
most_common_cylinder_layout_yeargroup_of_three: 0.00015903730273129931
```

Feature importance:

```
Feature importance in ordine decrescente:  
max_speed_km_per_h: 0.3713985690629088  
engine_hp: 0.16399304312874916  
maximum_torque_n_m: 0.09069403325993038  
number_of_gears: 0.035850168284121964  
rear_track_mm: 0.033207764992449505  
number_of_cylinders: 0.027423656713149846  
front_track_mm: 0.02661160078597661  
capacity_cm3: 0.024390063028885082  
curb_weight_kg: 0.024202116474767044  
width_mm: 0.02242857327101484  
valves_per_cylinder: 0.021480033246295403  
cylinder_layout: 0.01735579659855715  
length_mm: 0.015483382205441298  
wheelbase_mm: 0.014538917993152578  
turnover_of_maximum_torque_rpm: 0.012722970654188854  
cylinder_bore_mm: 0.012533755551563028  
fuel_tank_capacity_l: 0.011508027823784678  
year_from: 0.010749700210567035
```

## Capitolo 6: AdaBoost

Il secondo modello di regressione analizzato sui quattro dataset è l'AdaBoost. Le prestazioni sono state valutate tramite *10-fold Cross-Validation*, testando tutte le combinazioni dei seguenti iperparametri:

- `n_estimators`: i valori testati sono 10, 20 e 30, e rappresentano il numero di “weak learners” utilizzati per costruire il modello di boosting.
- `learning_rate`: sono stati provati i valori 0.1, 0.01 e 0.001; questo parametro controlla di quanto viene aggiornato il peso di ogni weak learner durante l'addestramento.

Questa configurazione ha permesso di esplorare e ottimizzare le performance del modello, individuando la combinazione di iperparametri più efficace per ciascun dataset.

## Valutazioni dei dataset

-preprocessed.csv

```
Migliori parametri: {'learning_rate': 0.1, 'n_estimators': 30, 'random_state': 0}
Migliore score: -1.829204880593648
Risultati della ricerca a griglia:
mean_test_neg_mean_absolute_error: [-1.03263349 -0.99347464 -0.97441152 -1.03740813 -1.04028405 -1.03179159
-1.0366517 -1.03926592 -1.03802536]
mean_test_neg_mean_squared_error: [-2.07087792 -1.9141284 -1.82920488 -2.09465147 -2.0941568 -2.06651891
-2.08563436 -2.09485456 -2.09314825]
mean_test_max_error: [-10.76262809 -10.32836986 -10.18337401 -10.91251937 -10.77585663
-10.78274288 -10.78953603 -10.79841506 -10.80331479]
```

Feature importance:

```
Importanza delle feature:
max_speed_km_per_h : 0.8744482465561646
engine_hp : 0.09349169039464952
year_to : 0.009093427132181319
front_suspension : 0.005687001199809287
engine_type : 0.005474325793626104
engine_hp_rpm : 0.004570557130535968
highway_fuel_per_100km_l : 0.003118253894453129
turnover_of_maximum_torque_rpm : 0.001559357038759476
width_mm : 0.001311522097050225
city_fuel_per_100km_l : 0.0011757142712788733
cylinder_bore_mm : 6.990449149135628e-05
```

-clustered\_plus\_prolog\_generated.csv

```
Migliori parametri: {'learning_rate': 0.1, 'n_estimators': 30, 'random_state': 0}
Migliore score: -1.2698369424293219
Risultati della ricerca a griglia:
mean_test_neg_mean_absolute_error: [-0.95392309 -0.87619283 -0.8389755 -1.01223532 -0.99620209 -0.9872308
-1.01530594 -1.01173399 -1.01573172]
mean_test_neg_mean_squared_error: [-1.6851891 -1.38309408 -1.26983694 -1.8991874 -1.86179754 -1.82042988
-1.90839928 -1.90680113 -1.92567109]
mean_test_max_error: [ -9.35525081 -8.85842704 -8.30931281 -10.12666858 -10.13156082
-9.87055951 -10.27103472 -10.18661934 -10.20302672]
```

Feature importance:

```
Importanza delle feature:
weight_to_power_ratio : 0.6781008871957722
max_speed_km_per_h : 0.2968071226548724
highway_fuel_per_100km_l : 0.0073871800229591365
length_power_ratio : 0.007346195210231009
specific_power : 0.00577643821504934
city_fuel_per_100km_l : 0.0031860404335125945
number_of_gears : 0.0013961362676032914
```

-prolog\_generated.csv

```
Migliori parametri: {'learning_rate': 0.1, 'n_estimators': 30, 'random_state': 0}
Migliore score: -1.78357893025848
Risultati della ricerca a griglia:
mean_test_neg_mean_absolute_error: [-1.03878759 -1.01017841 -1.00623262 -1.07140886 -1.06821771 -1.06612063
-1.07397676 -1.07227421 -1.07142799]
mean_test_neg_mean_squared_error: [-1.91040607 -1.7966276 -1.78357893 -2.03690223 -2.01915732 -2.01047636
-2.04862404 -2.03921617 -2.03314592]
mean_test_max_error: [ -9.73140213 -9.19237686 -9.32700059 -9.965322 -9.78454926
-9.78551687 -10.02892119 -9.98269474 -9.95179397]
```

Feature importance:

```
Importanza delle feature:
weight_to_power_ratio : 0.9355954128638821
length_power_ratio : 0.042844487620891415
specific_power : 0.00741892362046784
most_common_cylinder_layout_yeargroup_of_three : 0.006286019269592744
length_ratio : 0.004174712115397652
avg_trunk_capacity_brand : 0.0036804445097683705
```

-clustered.csv

```
Migliori parametri: {'learning_rate': 0.1, 'n_estimators': 30, 'random_state': 0}
Migliore score: -1.8309552710113954
Risultati della ricerca a griglia:
mean_test_neg_mean_absolute_error: [-1.03263349 -0.99690611 -0.97402657 -1.03740813 -1.04028405 -1.03179159
-1.0366517 -1.03926592 -1.03802536]
mean_test_neg_mean_squared_error: [-2.07087792 -1.92697538 -1.83095527 -2.09465147 -2.0941568 -2.06651891
-2.08563436 -2.09485456 -2.09314825]
mean_test_max_error: [-10.76262809 -10.36849557 -10.18774458 -10.91251937 -10.77585663
-10.78274288 -10.78953603 -10.79841506 -10.80331479]
```

Feature importance:

```
Importanza delle feature:
max_speed_km_per_h : 0.8744482465561644
engine_hp : 0.09349169039464143
year_to : 0.009093427132181642
front_suspension : 0.005756905691300799
engine_type : 0.005474325793634063
engine_hp_rpm : 0.004570557130535776
highway_fuel_per_100km_l : 0.0031182538944529922
turnover_of_maximum_torque_rpm : 0.0015593570387595144
width_mm : 0.0013115220970503752
city_fuel_per_100km_l : 0.001175714271278943
```

## Capitolo 7: Random Forest

Il terzo modello di regressione analizzato sui quattro dataset è il Random Forest. Le prestazioni di questo modello sono state valutate utilizzando la *10-fold Cross-Validation*, testando tutte le combinazioni dei seguenti iperparametri:

- `n_estimators`: i valori testati sono 10, 20 e 30, che rappresentano il numero di alberi decisionali generati nel modello.
- `max_depth`: sono stati testati i valori 4, 6 e 8, indicando la profondità massima consentita per ciascun albero decisionale.
- `min_samples_split`: con valori di 2, 4 e 6, questo parametro stabilisce il numero minimo di campioni necessari per suddividere un nodo interno.
- `min_samples_leaf`: i valori testati sono 1, 2 e 4, indicando il numero minimo di campioni richiesti per una foglia.
- `max_features`: i valori testati sono *sqrt* e *log2*. Con *sqrt*, il numero di feature selezionate per ciascun albero è la radice quadrata del totale, mentre *log2* seleziona un numero di feature pari al logaritmo in base 2 del totale.

Per ciascun dataset, sono state selezionate e riportate le cinque migliori combinazioni di iperparametri, poiché presentare tutte e diciotto le configurazioni risulterebbe poco leggibile.

### Valutazioni dei dataset

-preprocessed.csv

```
Migliori parametri: {'max_depth': 8, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 6, 'n_estimators': 30, 'random_state': 0}
Migliore score: -1.3106660211063716
Risultati della ricerca a griglia:
```

```
mean_test_neg_mean_absolute_error: [-1.1820947 -1.13261126 -1.09668441 -1.1820947 -1.13261126 -1.09668441
-1.1820947 -1.13261126 -1.09668441 -1.18186373 -1.1324723 -1.09659178
-1.18186373 -1.1324723 -1.09659178 -1.18186373 -1.1324723 -1.09659178
-1.18193583 -1.13248239 -1.09666877 -1.18193583 -1.13248239 -1.09666877
-1.18193583 -1.13248239 -1.09666877 -1.22014454 -1.17139207 -1.13847613
-1.22140854 -1.17209998 -1.13934548 -1.22392254 -1.17442497 -1.14069267
-1.22133226 -1.1721731 -1.1394554 -1.22133226 -1.1721731 -1.1394554
-1.22389223 -1.17452535 -1.14081504 -1.22864391 -1.17678325 -1.14183325
-1.22864391 -1.17678325 -1.14183325 -1.22864391 -1.17678325 -1.14183325
-0.98536405 -0.93261155 -0.91514845 -0.98552759 -0.92820599 -0.91263483
-0.97578023 -0.92554129 -0.91113868 -0.97021463 -0.91984006 -0.90826081
-0.97021463 -0.91984006 -0.90826081 -0.95717343 -0.91647697 -0.90537068
-0.95795328 -0.91554157 -0.90601937 -0.95795328 -0.91554157 -0.90601937
-0.95795328 -0.91554157 -0.90601937 -0.97321115 -0.95385009 -0.94046334]
```

```
mean_test_neg_mean_squared_error: [-2.83898449 -2.58351732 -2.43193147 -2.83898449 -2.58351732 -2.43193147
-2.83898449 -2.58351732 -2.43193147 -2.83676372 -2.58222746 -2.43106511
-2.83676372 -2.58222746 -2.43106511 -2.83676372 -2.58222746 -2.43106511
-2.83721778 -2.58190436 -2.4311348 -2.83721778 -2.58190436 -2.4311348
-2.83721778 -2.58190436 -2.4311348 -2.81472488 -2.64252813 -2.50872577
-2.82659699 -2.65075586 -2.5150012 -2.85439693 -2.6665614 -2.52486067
-2.82653391 -2.65150397 -2.51575233 -2.82653391 -2.65150397 -2.51575233
-2.85436263 -2.66737176 -2.52563946 -2.85906607 -2.66628389 -2.52404533
-2.85906607 -2.66628389 -2.52404533 -2.85906607 -2.66628389 -2.52404533
-1.96402545 -1.7389868 -1.6628416 -1.95858437 -1.72179796 -1.65846514
-1.92415113 -1.72606843 -1.65971155 -1.90069627 -1.69916687 -1.64925505
-1.90069627 -1.69916687 -1.64925505 -1.8648582 -1.6979164 -1.64572641
-1.87142333 -1.69380778 -1.64472836 -1.87142333 -1.69380778 -1.64472836]
```

```
mean_test_max_error: [-13.1825386 -12.48489162 -12.23796168 -13.1825386 -12.48489162
-12.23796168 -13.1825386 -12.48489162 -12.23796168 -13.1825386
-12.48489162 -12.23796168 -13.1825386 -12.48489162 -12.23796168
-13.1825386 -12.48489162 -12.23796168 -13.1825386 -12.44321701
-12.20934114 -13.1825386 -12.44321701 -12.20934114 -13.1825386
-12.44321701 -12.20934114 -12.43138403 -12.3841176 -12.10503124
-12.36912897 -12.35281797 -12.0787806 -12.40143841 -12.36897268
-12.09888437 -12.36912897 -12.35522394 -12.08038458 -12.36912897
-12.35522394 -12.08038458 -12.40143841 -12.37137866 -12.10048836
-12.42614214 -12.39760603 -12.13555389 -12.42614214 -12.39760603
-12.13555389 -12.42614214 -12.39760603 -12.13555389 -11.02321116
-10.69964374 -10.36412583 -11.28008845 -10.69855844 -10.42242155
-11.1595372 -10.82153721 -10.52895131 -11.16809669 -10.72588181
-10.4882173 -11.16809669 -10.72588181 -10.4882173 -11.16209526]
```

Feature importance:

```
Importanza delle feature:
max_speed_km_per_h : 0.31742819700748054
engine_hp : 0.22022606950313028
maximum_torque_n_m : 0.07938422066037451
number_of_gears : 0.04550556363265879
capacity_cm3 : 0.041299912628863664
rear_track_mm : 0.026119991468353206
number_of_cylinders : 0.02312388376945096
engine_hp_rpm : 0.022807691287891684
valves_per_cylinder : 0.02238880846326793
turnover_of_maximum_torque_rpm : 0.01835643188440403
width_mm : 0.015771764851105367
height_mm : 0.015720204745531995
front_track_mm : 0.014391528760004005
cylinder_bore_mm : 0.013395828067494151
curb_weight_kg : 0.009389928598057982
year_to : 0.009213281566370989
city_fuel_per_100km_l : 0.008592243073769787
```



## -clustered\_plus\_prolog\_generated.csv

```
Migliori parametri: {'max_depth': 8, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 6, 'n_estimators': 30, 'random_state': 0}  
Migliore score: -0.9474015080829933
```

Risultati della ricerca a griglia:

```
mean_test_neg_mean_absolute_error: [-0.92290439 -0.9005501 -0.88969071 -0.92290439 -0.9005501 -0.88969071  
-0.92290439 -0.9005501 -0.88969071 -0.92310443 -0.9006677 -0.8897691  
-0.92310443 -0.9006677 -0.8897691 -0.92310443 -0.9006677 -0.8897691  
-0.92310523 -0.90068107 -0.8898305 -0.92310523 -0.90068107 -0.8898305  
-0.92310523 -0.90068107 -0.8898305 -0.97936965 -0.99163823 -0.98414056  
-0.97936965 -0.99163823 -0.98414056 -0.97936965 -0.99163823 -0.98414056  
-0.97698267 -0.99062881 -0.98335967 -0.97698267 -0.99062881 -0.98335967  
-0.97698267 -0.99062881 -0.98335967 -0.97729567 -0.99077746 -0.98337097  
-0.97729567 -0.99077746 -0.98337097 -0.97729567 -0.99077746 -0.98337097  
-0.77381895 -0.76467928 -0.75177793 -0.77949625 -0.76465064 -0.75436988  
-0.77856668 -0.76256394 -0.75223342 -0.77819948 -0.76429106 -0.7537303  
-0.77819948 -0.76429106 -0.7537303 -0.777751 -0.76288363 -0.75128365  
-0.76796513 -0.7621652 -0.74933831 -0.76796513 -0.7621652 -0.74933831
```

```
mean_test_neg_mean_squared_error: [-1.72829004 -1.64490514 -1.58257114 -1.72829004 -1.64490514 -1.58257114  
-1.72829004 -1.64490514 -1.58257114 -1.73003755 -1.64580126 -1.58312456  
-1.73003755 -1.64580126 -1.58312456 -1.73003755 -1.64580126 -1.58312456  
-1.72997862 -1.64575349 -1.58354742 -1.72997862 -1.64575349 -1.58354742  
-1.72997862 -1.64575349 -1.58354742 -2.02044782 -2.00919901 -1.96846705  
-2.02044782 -2.00919901 -1.96846705 -2.02044782 -2.00919901 -1.96846705  
-2.00782509 -2.00272668 -1.96392508 -2.00782509 -2.00272668 -1.96392508  
-2.00782509 -2.00272668 -1.96392508 -2.00824398 -2.00285324 -1.96333968  
-2.00824398 -2.00285324 -1.96333968 -2.00824398 -2.00285324 -1.96333968  
-1.20849736 -1.17883005 -1.14470958 -1.21446373 -1.17743539 -1.15474965  
-1.21439269 -1.16508909 -1.14396519 -1.20936644 -1.1735788 -1.1507521  
-1.20936644 -1.1735788 -1.1507521 -1.21315683 -1.16695374 -1.1424056  
-1.18117087 -1.15911273 -1.13024875 -1.18117087 -1.15911273 -1.13024875
```

```
mean_test_max_error: [-11.07524972 -10.79214209 -10.50852825 -11.07524972 -10.79214209  
-10.50852825 -11.07524972 -10.79214209 -10.50852825 -11.09326678  
-10.80115062 -10.51453394 -11.09326678 -10.80115062 -10.51453394  
-11.09326678 -10.80115062 -10.51453394 -11.09326678 -10.80115062  
-10.51453394 -11.09326678 -10.80115062 -10.51453394 -11.09326678  
-10.80115062 -10.51453394 -11.88874221 -11.63185719 -11.52894328  
-11.88874221 -11.63185719 -11.52894328 -11.88874221 -11.63185719  
-11.52894328 -11.88673076 -11.63085146 -11.52827279 -11.88673076  
-11.63085146 -11.52827279 -11.88673076 -11.63085146 -11.52827279  
-11.88858678 -11.63885239 -11.52534489 -11.88858678 -11.63885239  
-11.52534489 -11.88858678 -11.63885239 -11.52534489 -9.23837149  
-8.89400549 -9.03834211 -9.24205447 -8.8787274 -9.0333587  
-9.12034391 -8.79349602 -8.90713229 -9.3778244 -8.94036017  
-9.07128958 -9.3778244 -8.94036017 -9.07128958 -9.25655937
```

## Feature importance:

```
Importanza delle feature:
weight_to_power_ratio : 0.20694746531998626
max_speed_km_per_h : 0.1908592232343832
length_power_ratio : 0.10047876373850186
engine_hp : 0.09503718197765383
specific_power : 0.06089165888498991
max_speed_power_ratio : 0.05783812826461269
power : 0.05634429714029278
maximum_torque_n_m : 0.04744581824428263
number_of_gears : 0.02510722041390823
capacity_cm3 : 0.019525120369678038
valves_per_cylinder : 0.01948671184965745
cylinder_bore_mm : 0.010679651034811
curb_weight_kg : 0.010547969845775404
number_of_cylinders : 0.010486013081845193
height_mm : 0.010437790896041518
length_ratio : 0.007729890119955189
rear_track_mm : 0.007611505764136547
city_fuel_per_100km_l : 0.004930094963284217
width_mm : 0.004527563517719855
drag_coefficient : 0.004100040734681906
turnover_of_maximum_torque_rpm : 0.003986730794277463
year_from : 0.0033357539712348143
front_track_mm : 0.0032599268581305113
engine_hp_rpm : 0.0031004624091515677
year_to : 0.002853299230561402
cluster : 0.002790950387534164
```

## -prolog\_generated.csv

```
Migliori parametri: {'max_depth': 8, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 6, 'n_estimators': 30, 'random_state': 0}
Migliore score: -1.1889522039611573
Risultati della ricerca a griglia:
mean_test_neg_mean_absolute_error: [-0.96053023 -0.93603496 -0.92950504 -0.96053023 -0.93603496 -0.92950504
-0.96230393 -0.93668771 -0.92991195 -0.96082938 -0.93567012 -0.92932455
-0.96082938 -0.93567012 -0.92932455 -0.96260308 -0.93632287 -0.92973146
-0.9625278 -0.93636472 -0.92972189 -0.9625278 -0.93636472 -0.92972189
-0.9625278 -0.93636472 -0.92972189 -0.96053023 -0.93603496 -0.92950504
-0.96053023 -0.93603496 -0.92950504 -0.96230393 -0.93668771 -0.92991195
-0.96082938 -0.93567012 -0.92932455 -0.96082938 -0.93567012 -0.92932455
-0.96260308 -0.93632287 -0.92973146 -0.9625278 -0.93636472 -0.92972189
-0.9625278 -0.93636472 -0.92972189 -0.9625278 -0.93636472 -0.92972189
-0.84590478 -0.84089298 -0.83574497 -0.84826217 -0.84281987 -0.83819243
-0.8441255 -0.83666839 -0.8350819 -0.84639328 -0.84079862 -0.83736971]
```

```
mean_test_neg_mean_squared_error: [-1.77606871 -1.68915601 -1.67204896 -1.77606871 -1.68915601 -1.67204896
-1.7844315 -1.6926064 -1.67440901 -1.77789341 -1.68784906 -1.67117003
-1.77789341 -1.68784906 -1.67117003 -1.78625619 -1.69130035 -1.67353049
-1.78601288 -1.69158542 -1.67355888 -1.78601288 -1.69158542 -1.67355888
-1.78601288 -1.69158542 -1.67355888 -1.77606871 -1.68915601 -1.67204896
-1.77606871 -1.68915601 -1.67204896 -1.7844315 -1.6926064 -1.67440901
-1.77789341 -1.68784906 -1.67117003 -1.77789341 -1.68784906 -1.67117003
-1.78625619 -1.69130035 -1.67353049 -1.78601288 -1.69158542 -1.67355888
-1.78601288 -1.69158542 -1.67355888 -1.78601288 -1.69158542 -1.67355888
-1.38145174 -1.36634588 -1.35077158 -1.39276143 -1.37045721 -1.35873794
-1.37886632 -1.34810264 -1.34431687 -1.38703542 -1.36439282 -1.35518359]
```



```
mean_test_max_error: [-10.75568278 -10.5675303 -10.58069147 -10.75568278 -10.5675303
-10.58069147 -10.77235187 -10.57586485 -10.58624783 -10.75265612
-10.54142941 -10.56329087 -10.75265612 -10.54142941 -10.56329087
-10.76932522 -10.54976396 -10.56884723 -10.76951752 -10.53239607
-10.56349087 -10.76951752 -10.53239607 -10.56349087 -10.76951752
-10.53239607 -10.56349087 -10.75568278 -10.5675303 -10.58069147
-10.75568278 -10.5675303 -10.58069147 -10.77235187 -10.57586485
-10.58624783 -10.75265612 -10.54142941 -10.56329087 -10.75265612
-10.54142941 -10.56329087 -10.76932522 -10.54976396 -10.56884723
-10.76951752 -10.53239607 -10.56349087 -10.76951752 -10.53239607
-10.56349087 -10.76951752 -10.53239607 -10.56349087 -9.40514241
-9.31720896 -9.19539499 -9.47549339 -9.22481906 -9.12532432
-9.33644142 -9.12887966 -8.97207483 -9.41042021 -9.16778575]
```

Feature importance:

```
Importanza delle feature:
weight_to_power_ratio : 0.3809143309977849
length_power_ratio : 0.25293531808823366
specific_power : 0.1141632045427596
max_speed_power_ratio : 0.09930883651295189
power : 0.07453327516994976
length_ratio : 0.02786713854405837
drag_coefficient : 0.016591058682757943
number_cylinder_capacity_ratio : 0.014481169269079781
avg_trunk_capacity_brand : 0.007280322673268886
most_powerful_car_brand : 0.005225411252857861
most_common_cylinder_layout_yeargroup_of_three : 0.0036648951088540196
shape : 0.0030350391574433715
```

-clustered.csv

```
Migliori parametri: {'max_depth': 8, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 6, 'n_estimators': 30, 'random_state': 0}
Migliore score: -1.3455442365792218
Risultati della ricerca a griglia:
mean_test_neg_mean_absolute_error: [-1.20018844 -1.16259048 -1.15155097 -1.20018844 -1.16095756 -1.15057369
-1.20018844 -1.16095756 -1.15057369 -1.20044954 -1.16142008 -1.15087233
-1.20044954 -1.16142008 -1.15087233 -1.20044954 -1.16142008 -1.15087233
-1.20035693 -1.16325433 -1.15194654 -1.20035693 -1.16325433 -1.15194654
-1.20035693 -1.16325433 -1.15194654 -1.26862351 -1.20760842 -1.19626594
-1.26862351 -1.20760842 -1.19626594 -1.26862351 -1.20760842 -1.19626594
-1.27052489 -1.20847843 -1.19693062 -1.27052489 -1.20847843 -1.19693062
-1.27052489 -1.20847843 -1.19693062 -1.27065618 -1.2085158 -1.19688989
-1.27065618 -1.2085158 -1.19688989 -1.27065618 -1.2085158 -1.19688989
-0.97137393 -0.94416301 -0.94310019 -0.98394798 -0.95165608 -0.94380166
-0.98272413 -0.95102044 -0.94548775 -0.98472102 -0.95898986 -0.94964908
-0.98472102 -0.95898986 -0.94964908 -0.98297774 -0.95564425 -0.94910874
-0.98015341 -0.95158037 -0.94435028 -0.98015341 -0.95158037 -0.94435028]
```

```
mean_test_neg_mean_squared_error: [-2.80206166 -2.66333663 -2.61604412 -2.80206166 -2.65722568 -2.6134482
-2.80206166 -2.65722568 -2.6134482 -2.80271999 -2.65899461 -2.61489508
-2.80271999 -2.65899461 -2.61489508 -2.80271999 -2.65899461 -2.61489508
-2.80250608 -2.66781847 -2.61848986 -2.80250608 -2.66781847 -2.61848986
-2.80250608 -2.66781847 -2.61848986 -3.16830078 -2.9008254 -2.83368703
-3.16830078 -2.9008254 -2.83368703 -3.16830078 -2.9008254 -2.83368703
-3.17147777 -2.90214242 -2.83450876 -3.17147777 -2.90214242 -2.83450876
-3.17147777 -2.90214242 -2.83450876 -3.17366129 -2.90299628 -2.8346369
-3.17366129 -2.90299628 -2.8346369 -3.17366129 -2.90299628 -2.8346369
-1.90626278 -1.80865818 -1.80841177 -1.9567885 -1.83286674 -1.81323243
-1.9656416 -1.8324339 -1.82516903 -1.95767888 -1.84752366 -1.82369915
-1.95767888 -1.84752366 -1.82369915 -1.9712278 -1.84739679 -1.8361148
```

```
mean_test_max_error: [-12.48633323 -12.39054945 -12.40002764 -12.48633323 -12.36308001
-12.38171468 -12.48633323 -12.36308001 -12.38171468 -12.48647436
-12.36318592 -12.38178529 -12.48647436 -12.36318592 -12.38178529
-12.48647436 -12.36318592 -12.38178529 -12.48647436 -12.39783583
-12.39184523 -12.48647436 -12.39783583 -12.39184523 -12.48647436
-12.39783583 -12.39184523 -12.70147445 -12.47923376 -12.50404418
-12.70147445 -12.47923376 -12.50404418 -12.70147445 -12.47923376
-12.50404418 -12.69331095 -12.47515201 -12.50132301 -12.69331095
-12.47515201 -12.50132301 -12.69331095 -12.47515201 -12.50132301
-12.69582149 -12.47729277 -12.50275019 -12.69582149 -12.47729277
-12.50275019 -12.69582149 -12.47729277 -12.50275019 -11.08000868
-10.748719 -10.73962664 -11.23764173 -10.76456234 -10.73320588
-11.24650365 -10.74046173 -10.75356258 -11.19562112 -10.72816355
-10.68755768 -11.19562112 -10.72816355 -10.68755768 -11.25832984
-10.75918867 -10.75575382 -11.14872394 -10.55264931 -10.54639757
-11.14872394 -10.55264931 -10.54639757 -11.14872394 -10.55264931
```

## Feature Importance:

```
Importanza delle feature:
max_speed_km_per_h : 0.24883291395425194
engine_hp : 0.17847789459561386
maximum_torque_nm : 0.14503558541059688
cluster : 0.05015493534380964
capacity_cm3 : 0.048528471191440214
number_of_gears : 0.04027920448863228
number_of_cylinders : 0.027894422060019502
rear_track_mm : 0.025242279179268374
turnover_of_maximum_torque_rpm : 0.019908468291757023
front_track_mm : 0.019884461633035748
engine_hp_rpm : 0.018948520378024566
valves_per_cylinder : 0.01791664040086989
curb_weight_kg : 0.01753813936028917
height_mm : 0.016940378512392957
cylinder_layout : 0.013676114393402694
highway_fuel_per_100km_l : 0.009181341419608652
fuel_grade : 0.007516961140794167
year_to : 0.007482207883960614
wheelbase_mm : 0.00715049384764939
cylinder_bore_mm : 0.006879331051118756
engine_type : 0.0065615089390504
fuel_tank_capacity_l : 0.006552404225200338
year_from : 0.006462305102659037
```

## Capitolo 8: Alberi di Decisione

Il quarto modello di regressione analizzato sui quattro dataset è il Decision Tree. Le prestazioni di questo modello sono state valutate attraverso la *10-fold Cross-Validation*, testando tutte le combinazioni dei seguenti iperparametri:

- `max_depth`: i valori testati sono *None* e un intervallo da 3 a 8, che rappresentano la profondità massima consentita per l'albero.
- `min_samples_split`: con valori di 2, 5 e 10, questo parametro indica il numero minimo di campioni necessari per suddividere un nodo interno.
- `min_samples_leaf`: i valori testati sono 1, 2 e 4, indicando il numero minimo di campioni richiesti per una foglia dell'albero.
- `max_features`: sono stati provati i valori 0.5, *sqrt* e *log2*. Con 0.5, il modello considera circa il 50% delle feature per ciascuna suddivisione; *sqrt* seleziona il numero di feature pari alla radice quadrata del totale, e *log2* seleziona il numero di feature pari al logaritmo in base 2 del totale.

### Valutazioni dei dataset

-preprocessed.csv

```
Migliori parametri: {'max_depth': 8, 'max_features': 0.5, 'min_samples_leaf': 4, 'min_samples_split': 10, 'random_state': 0}
Risultati della ricerca a griglia:
mean_test_neg_mean_absolute_error: [-0.95406852 -0.97288869 -0.95403704 -0.94687102 -0.94165898 -0.92248226
-0.90715303 -0.90715303 -0.88785373 -1.1101037  -1.07788907 -1.06693443
-1.0778289  -1.0762162  -1.10139989 -1.04436821 -1.04436821 -1.07457401
-1.11542668 -1.10787909 -1.09040887 -1.0614416  -1.10342075 -1.0482794
-1.0696485  -1.0696485  -1.06068911 -1.23772171 -1.23772171 -1.23772171
-1.23772171 -1.23772171 -1.23772171 -1.23772171 -1.23772171 -1.23772171
-1.98376792 -1.98376792 -1.98376792 -1.98376792 -1.98376792 -1.98376792
-1.98376792 -1.98376792 -1.98376792 -1.56124837 -1.56124837 -1.56124837
-1.56124837 -1.56124837 -1.56124837 -1.56124837 -1.56124837 -1.56124837
-1.1827588  -1.1827588  -1.1827588  -1.1827588  -1.1827588  -1.1827588
-1.1827588  -1.1827588  -1.1827588  -1.69603662 -1.69603662 -1.69603662]
```

```

mean_test_neg_mean_squared_error: [-2.01954228 -2.03109743 -1.97000266 -1.86138669 -1.86170089 -1.85546913
-1.71058178 -1.71058178 -1.66568083 -2.65417153 -2.57577531 -2.45089885
-2.48328852 -2.33543699 -2.48381688 -2.32778956 -2.32778956 -2.47133021
-2.60510338 -2.78563842 -2.5564629 -2.34349913 -2.53522237 -2.28015224
-2.34546725 -2.34546725 -2.41239894 -2.85847812 -2.85847812 -2.85847812
-2.85847812 -2.85847812 -2.85847812 -2.85847812 -2.85847812
-6.91234406 -6.91234406 -6.91234406 -6.91234406 -6.91234406 -6.91234406
-6.91234406 -6.91234406 -6.91234406 -4.58788352 -4.58788352 -4.58788352
-4.58788352 -4.58788352 -4.58788352 -4.58788352 -4.58788352
-2.67085612 -2.67085612 -2.67085612 -2.67085612 -2.67085612 -2.67085612
-2.67085612 -2.67085612 -2.67085612 -5.18991712 -5.18991712 -5.18991712
-5.1649788 -5.1649788 -5.1649788 -5.1649788 -5.1649788 -5.1649788

mean_test_max_error: [-11.4 -13.12666667 -11.79583333 -10.36 -10.4825
-10.96709524 -10.34909524 -10.34909524 -10.15687302 -13.18
-16.2175 -12.33490079 -11.3935 -10.76083333 -11.24475
-13.02792857 -13.02792857 -12.00338889 -11.77 -15.07916667
-13.11952381 -10.55 -10.735 -10.27751984 -11.50180952
-11.50180952 -12.41166667 -11.96701256 -11.96701256 -11.96701256
-11.96701256 -11.96701256 -11.96701256 -11.96701256 -11.96701256
-14.71424142 -14.71424142 -14.71424142 -14.71424142 -14.71424142
-14.71424142 -14.71424142 -14.71424142 -14.71424142 -14.71424142
-13.22567369 -13.22567369 -13.22567369 -13.22567369 -13.22567369
-13.22567369 -13.22567369 -13.22567369 -13.22567369 -11.30593275
-11.30593275 -11.30593275 -11.30593275 -11.30593275 -11.30593275
-11.30593275 -11.30593275 -11.30593275 -14.92761169 -14.92761169

```

Feature importance:

```

Importanza delle feature:
engine_hp : 0.6752494094661372
max_speed_km_per_h : 0.18134542318496272
maximum_torque_n_m : 0.0611072672643531
curb_weight_kg : 0.024203466110284726
height_mm : 0.010355424117589688
fuel_tank_capacity_l : 0.008516387241955468
capacity_cm3 : 0.007322705516901904
wheelbase_mm : 0.0053377605067307056
transmission : 0.00433305569914654
cylinder_bore_mm : 0.0033350573446861392
year_from : 0.002362870981554085
length_mm : 0.0022469610026770617
width_mm : 0.002017363623029572
number_of_gears : 0.0014013008464739327
city_fuel_per_100km_l : 0.0013968745855186824
engine_hp_rpm : 0.0013824095695771615
stroke_cycle_mm : 0.0013018858004572808
highway_fuel_per_100km_l : 0.0008796711254293002
max_trunk_capacity_l : 0.0007223425531201795
turnover_of_maximum_torque_rpm : 0.0006079362984844077
back_suspension : 0.0005959382484804993
minimum_trunk_capacity_l : 0.0005935106223893565
injection_type : 0.0005611754982151075

```

## -clustered\_plus\_prolog\_generated.csv

Migliori parametri: {'max\_depth': 8, 'max\_features': 0.5, 'min\_samples\_leaf': 4, 'min\_samples\_split': 10, 'random\_state': 0}  
Risultati della ricerca a griglia:

```
mean_test_neg_mean_absolute_error: [-0.87082686 -0.86746696 -0.85341014 -0.85855391 -0.85212433 -0.83460488
-0.82459653 -0.82459653 -0.83393199 -0.99112334 -0.96205231 -0.93075894
-0.95198105 -0.97757746 -0.94569061 -0.91792172 -0.91792172 -0.88114396
-1.01242577 -1.01786105 -1.04957306 -0.98479568 -0.98437343 -0.97621225
-0.96836259 -0.96836259 -0.9710625 -1.23067968 -1.23067968 -1.23067968
-1.23067968 -1.23067968 -1.23067968 -1.23067968 -1.23067968 -1.23067968
-1.62528291 -1.62528291 -1.62528291 -1.62528291 -1.62528291 -1.62528291
-1.62528291 -1.62528291 -1.62528291 -1.70288975 -1.70288975 -1.70288975
-1.70288975 -1.70288975 -1.70288975 -1.70288975 -1.70288975 -1.70288975
-0.99220902 -0.99220902 -0.99220902 -0.99220902 -0.99220902 -0.99220902
-0.99220902 -0.99220902 -0.99220902 -1.44740271 -1.44740271 -1.44740271
```

```
mean_test_neg_mean_squared_error: [-1.56584434 -1.54991604 -1.48932091 -1.60797892 -1.49011335 -1.46537906
-1.42287358 -1.42287358 -1.44427769 -2.20824323 -1.94067308 -1.85321694
-1.91708567 -2.146453 -1.87938641 -1.70903426 -1.70903426 -1.58530328
-2.19368663 -2.14762103 -2.30945651 -2.10552001 -1.98876013 -1.99180049
-1.98631474 -1.98631474 -2.02044864 -2.6855745 -2.6855745 -2.6855745
-2.6855745 -2.6855745 -2.6855745 -2.6855745 -2.6855745 -2.6855745
-4.66763814 -4.66763814 -4.66763814 -4.66763814 -4.66763814 -4.66763814
-4.66763814 -4.66763814 -4.66763814 -5.28898702 -5.28898702 -5.28898702
-5.28898702 -5.28898702 -5.28898702 -5.28898702 -5.28898702 -5.28898702
-1.896916 -1.896916 -1.896916 -1.896916 -1.896916 -1.896916
-1.896916 -1.896916 -1.896916 -4.02435512 -4.02435512 -4.02435512
-4.02435512 -4.02435512 -4.02435512 -4.02435512 -4.02435512 -4.02435512
-3.7674132 -3.7674132 -3.7674132 -3.7674132 -3.7674132 -3.7674132
```

```
mean_test_max_error: [ -9.13 -8.86666667 -9.40805556 -9.56166667 -9.4325
-10.12366667 -8.7655 -8.7655 -8.96205952 -12.59
-10.50083333 -10.56119444 -9.42666667 -10.37583333 -11.26107143
-9.58807143 -9.58807143 -8.21452778 -11.7 -9.6725
-10.21222222 -10.01833333 -10.13 -10.66958333 -10.1075
-10.1075 -10.63869048 -10.66764509 -10.66764509 -10.66764509
-10.66764509 -10.66764509 -10.66764509 -10.66764509 -10.66764509
-10.66764509 -14.53809327 -14.53809327 -14.53809327 -14.53809327
-14.53809327 -14.53809327 -14.53809327 -14.53809327 -14.53809327
-14.90423262 -14.90423262 -14.90423262 -14.90423262 -14.90423262
-14.90423262 -14.90423262 -14.90423262 -14.90423262 -10.17795582
-10.17795582 -10.17795582 -10.17795582 -10.17795582 -10.17795582
-10.17795582 -10.17795582 -10.17795582 -12.85537945 -12.85537945
-12.85537945 -12.85537945 -12.85537945 -12.85537945 -12.85537945
```



## Feature importance:

```
max_speed_power_ratio : 0.37225895544233556
max_speed_km_per_h : 0.35326620508255163
length_power_ratio : 0.121337831145888
weight_to_power_ratio : 0.08352694657434492
engine_type : 0.013485041408930464
engine_hp : 0.009539852234960549
maximum_torque_n_m : 0.009375671043219572
number_of_gears : 0.005348460656675702
specific_power : 0.005070671101521841
transmission : 0.00425849478674309
year_from : 0.002887922766428792
fuel_tank_capacity_l : 0.002678485724259412
city_fuel_per_100km_l : 0.001907695097259441
front_track_mm : 0.0018734060602488658
curb_weight_kg : 0.0016736793874907823
max_trunk_capacity_l : 0.0014254366752169184
capacity_cm3 : 0.0013891038354783487
height_mm : 0.0013569644389480472
back_suspension : 0.0012144509551590943
wheelbase_mm : 0.0009715580461990366
```

## -prolog\_generated.csv

```
Migliori parametri: {'max_depth': 8, 'max_features': 0.5, 'min_samples_leaf': 2, 'min_samples_split': 5, 'random_state': 0}
Risultati della ricerca a griglia:
```

```
mean_test_neg_mean_absolute_error: [-1.01626131 -0.98972164 -0.96162766 -0.96977864 -0.97149648 -0.95418193
 -0.92424098 -0.92424098 -0.9237375 -1.04958124 -1.03105337 -0.98538622
 -1.04758709 -0.98066183 -0.97195092 -0.98331899 -0.98331899 -0.9460989
 -1.04958124 -1.03105337 -0.98538622 -1.04758709 -0.98066183 -0.97195092
 -0.98331899 -0.98331899 -0.9460989 -1.24119608 -1.24119608 -1.24119608
 -1.24119608 -1.24119608 -1.24119608 -1.24119608 -1.24119608 -1.24119608
 -1.55410624 -1.55410624 -1.55410624 -1.55093164 -1.55093164 -1.55093164
 -1.55093164 -1.55093164 -1.55093164 -1.55093164 -1.55410624 -1.55410624
 -1.55093164 -1.55093164 -1.55093164 -1.55093164 -1.55093164 -1.55093164
 -1.07224412 -1.07224412 -1.07224412 -1.07224412 -1.07224412 -1.07224412
 -1.07224412 -1.07224412 -1.07224412 -1.25407646 -1.25407646 -1.25407646
 -1.25994892 -1.25994892 -1.25994892 -1.25994892 -1.25994892 -1.25994892]
```

```
mean_test_neg_mean_squared_error: [-2.07785972 -1.95262594 -1.87740632 -1.85791878 -1.84307951 -1.8924117
 -1.6561584 -1.6561584 -1.6376452 -2.29513862 -2.34140083 -2.01306035
 -2.28923521 -1.92148855 -1.93944771 -1.96013936 -1.96013936 -1.79889902
 -2.29513862 -2.34140083 -2.01306035 -2.28923521 -1.92148855 -1.93944771
 -1.96013936 -1.96013936 -1.79889902 -2.82135353 -2.82135353 -2.82135353
 -2.82135353 -2.82135353 -2.82135353 -2.82135353 -2.82135353 -2.82135353
 -4.57294477 -4.57294477 -4.57294477 -4.47908396 -4.47908396 -4.47908396
 -4.47908396 -4.47908396 -4.47908396 -4.57294477 -4.57294477 -4.57294477
 -4.47908396 -4.47908396 -4.47908396 -4.47908396 -4.47908396 -4.47908396
 -2.10672366 -2.10672366 -2.10672366 -2.10672366 -2.10672366 -2.10672366
 -2.10672366 -2.10672366 -2.10672366 -2.9506941 -2.9506941 -2.9506941
 -2.99275647 -2.99275647 -2.99275647 -2.99275647 -2.99275647 -2.99275647
 -2.9506941 -2.9506941 -2.9506941 -2.99275647 -2.99275647 -2.99275647]
```

```
mean_test_max_error: [-10.14      -9.77583333 -9.70818651 -9.57      -8.5075
-10.99191667 -8.39616667 -8.39616667 -9.18211508 -13.15
-12.43833333 -11.21617063 -10.29833333 -9.77166667 -9.71008333
-9.41583333 -9.41583333 -9.39110606 -13.15      -12.43833333
-11.21617063 -10.29833333 -9.77166667 -9.71008333 -9.41583333
-9.41583333 -9.39110606 -11.82155957 -11.82155957 -11.82155957
-11.82155957 -11.82155957 -11.82155957 -11.82155957 -11.82155957
-11.82155957 -15.20121934 -15.20121934 -15.20121934 -14.45653001
-14.45653001 -14.45653001 -14.45653001 -14.45653001 -14.45653001
-15.20121934 -15.20121934 -15.20121934 -14.45653001 -14.45653001
-14.45653001 -14.45653001 -14.45653001 -14.45653001 -9.43257182
-9.43257182 -9.43257182 -9.43257182 -9.43257182 -9.43257182]
```

Feature importance:

```
Importanza delle feature:
max_speed_power_ratio : 0.4309311541108177
weight_to_power_ratio : 0.32357746964689554
specific_power : 0.12583564816265733
length_power_ratio : 0.09208031123657731
drag_coefficient : 0.0071950489921374025
length_ratio : 0.006640492750786971
most_powerful_car_brand : 0.003532717730352478
avg_trunk_capacity_brand : 0.0028161241323684874
most_common_cylinder_layout_yeargroup_of_three : 0.0026305725777830877
shape : 0.0021684105359491427
power : 0.0019199878316438308
number_cylinder_capacity_ratio : 0.0006720622920305918
```

- clustered.csv

```
Migliori parametri: {'max_depth': 8, 'max_features': 0.5, 'min_samples_leaf': 2, 'min_samples_split': 5, 'random_state': 0}
Risultati della ricerca a griglia:
mean_test_neg_mean_absolute_error: [-0.98998767 -0.93906989 -0.92987158 -0.93304408 -0.96705314 -0.92785665
-0.91801374 -0.91801374 -0.91578313 -1.09207081 -1.04230151 -1.06001868
-1.04547505 -1.03679788 -1.08075251 -1.05851952 -1.05851952 -1.03803201
-1.13404249 -1.10655193 -1.05103486 -1.09240364 -1.11846998 -1.05816345
-1.07451745 -1.07451745 -1.04877331 -1.33440206 -1.33440206 -1.33440206
-1.33440206 -1.33440206 -1.33440206 -1.33440206 -1.33440206 -1.33440206
-1.6022303 -1.6022303 -1.6022303 -1.6022303 -1.6022303 -1.6022303
-1.6022303 -1.6022303 -1.6022303 -1.76363359 -1.76363359 -1.76363359
-1.76363359 -1.76363359 -1.76363359 -1.76363359 -1.76363359 -1.76363359
-1.21373643 -1.21373643 -1.21373643 -1.21373643 -1.21373643 -1.21373643
-1.21373643 -1.21373643 -1.21373643 -1.54200932 -1.54200932 -1.54200932
-1.54200932 -1.54200932 -1.54200932 -1.54200932 -1.54200932 -1.54200932
-1.46922774 -1.46922774 -1.46922774 -1.46922774 -1.46922774 -1.46922774]
```

```
mean_test_neg_mean_squared_error: [-2.20504698 -1.9002449 -1.88793785 -1.82565813 -2.09621867 -1.8245382
-1.74716035 -1.74716035 -1.75355087 -2.65172909 -2.29607238 -2.39659104
-2.2972371 -2.28480615 -2.58784494 -2.86931328 -2.86931328 -2.31262137
-2.92270436 -2.66155684 -2.45288867 -2.55487341 -2.78831216 -2.38670425
-2.63818329 -2.63818329 -2.3584401 -3.41403672 -3.41403672 -3.41403672
-3.41403672 -3.41403672 -3.41403672 -3.41403672 -3.41403672 -3.41403672
-4.82290805 -4.82290805 -4.82290805 -4.82290805 -4.82290805 -4.82290805
-4.82290805 -4.82290805 -4.82290805 -5.50024482 -5.50024482 -5.50024482
-5.50024482 -5.50024482 -5.50024482 -5.50024482 -5.50024482 -5.50024482
-2.82875861 -2.82875861 -2.82875861 -2.82875861 -2.82875861 -2.82875861
-2.82875861 -2.82875861 -2.82875861 -4.468922 -4.468922 -4.468922
-4.468922 -4.468922 -4.468922 -4.468922 -4.468922 -4.468922
-4.17143414 -4.17143414 -4.17143414 -4.17143414 -4.17143414 -4.17143414]
```

```
mean_test_max_error: [-12.73 -9.4975 -10.57603968 -11.22833333 -10.75666667
-9.68420635 -10.10721429 -10.10721429 -9.05716667 -14.71
-11.19833333 -12.31339286 -12.24166667 -11.99083333 -10.83672619
-11.49185897 -11.49185897 -10.89775794 -13.45 -12.49833333
-12.33163889 -11.22466667 -12.77083333 -10.51133333 -11.28519048
-11.28519048 -11.09878175 -12.53899934 -12.53899934 -12.53899934
-12.53899934 -12.53899934 -12.53899934 -12.53899934 -12.53899934
-12.53899934 -14.26503743 -14.26503743 -14.26503743 -14.26503743
-14.26503743 -14.26503743 -14.26503743 -14.26503743 -14.26503743
-13.13957691 -13.13957691 -13.13957691 -13.13957691 -13.13957691
-13.13957691 -13.13957691 -13.13957691 -13.13957691 -12.01585287
-12.01585287 -12.01585287 -12.01585287 -12.01585287 -12.01585287
-12.01585287 -12.01585287 -12.01585287 -13.63780352 -13.63780352]
```

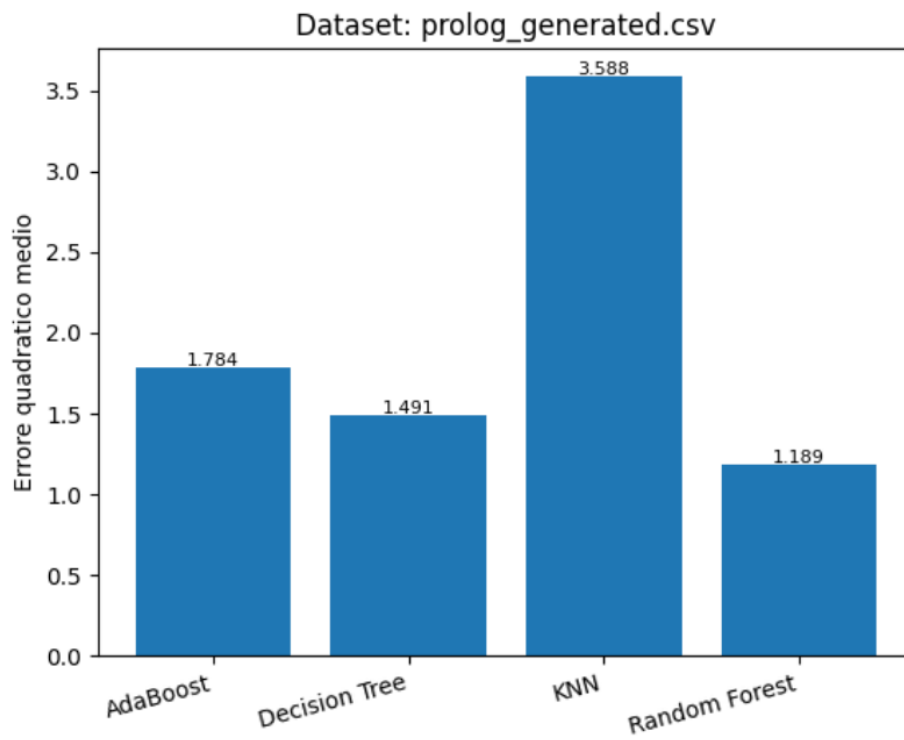
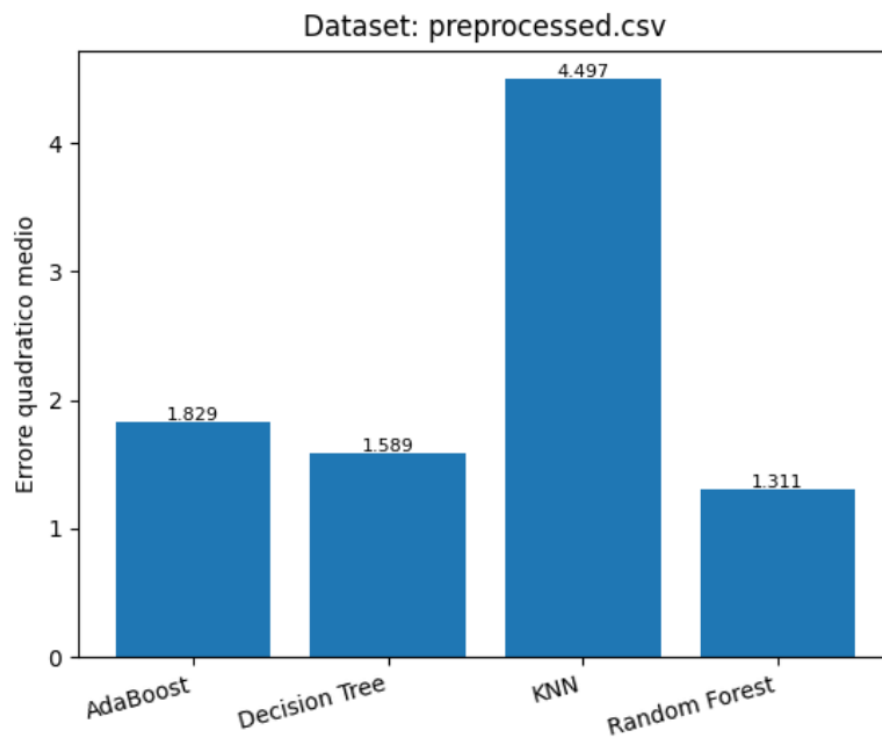
Feature importance:

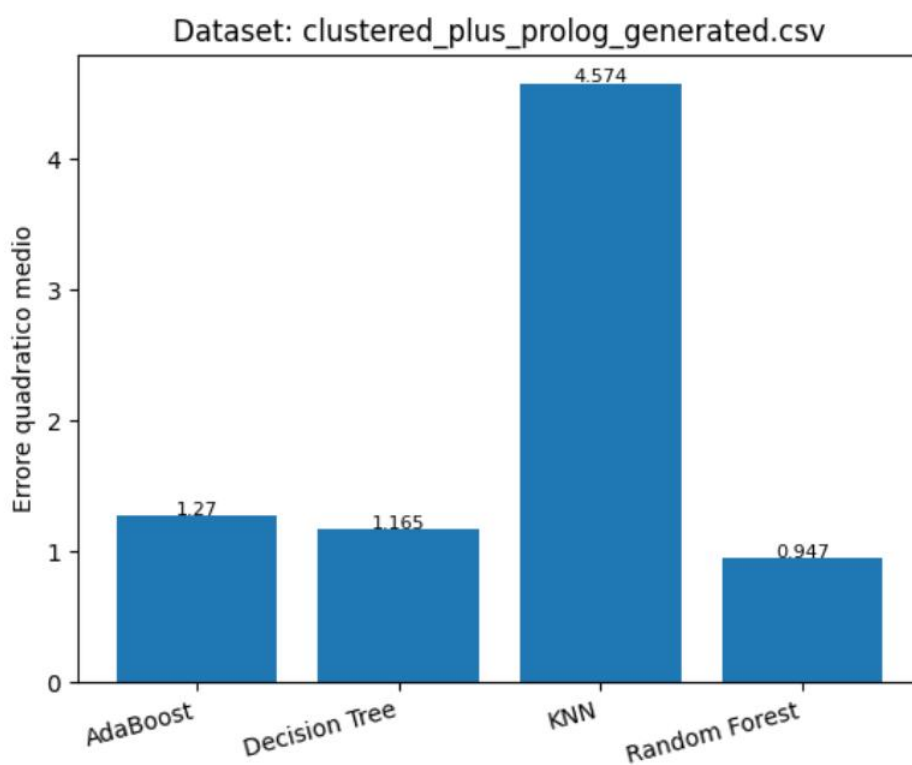
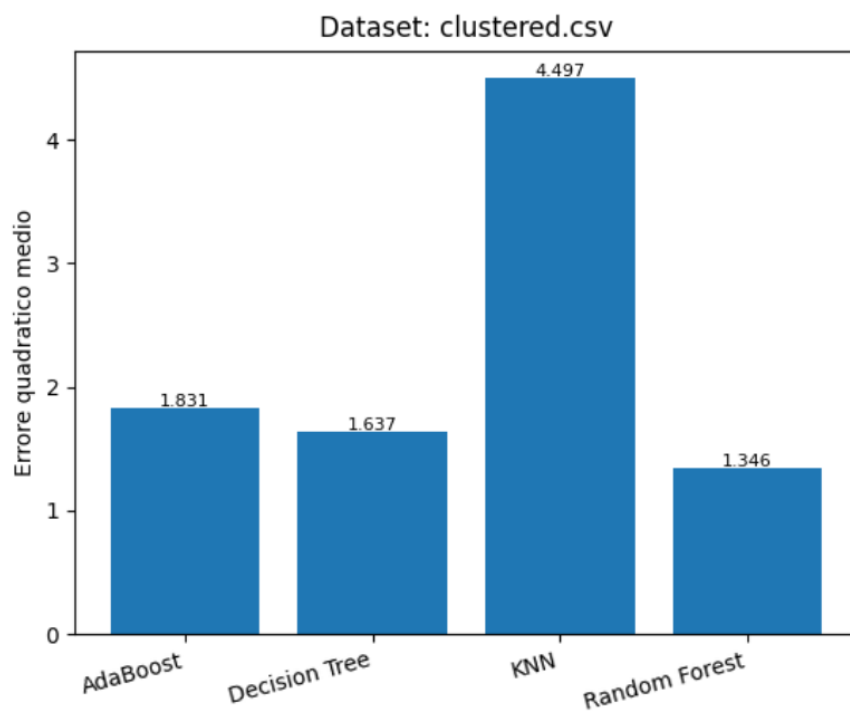
```
maximum_torque_n_m : 0.4709257463314697
max_speed_km_per_h : 0.2301589389950234
engine_hp : 0.07928654766982461
engine_hp_rpm : 0.07805873683528998
injection_type : 0.0521320704775404
fuel_tank_capacity_l : 0.03160318003930562
curb_weight_kg : 0.008901353409269789
transmission : 0.005566513369863605
cylinder_bore_mm : 0.005483402481890239
capacity_cm3 : 0.005462546066726138
engine_type : 0.00538359350256444
wheelbase_mm : 0.0035190678797091196
highway_fuel_per_100km_l : 0.0025789074776767225
model : 0.0024224574756088572
city_fuel_per_100km_l : 0.0023860906382428986
height_mm : 0.0023821250445104115
width_mm : 0.0021023440805896985
front_track_mm : 0.0019085455266343546
length_mm : 0.0017542929852301535
```



## Capitolo 9 : Risultati Finali

Possiamo confrontare adesso i vari risultati ottenuti confrontando tutti i modelli





Come possiamo vedere dai risultati ottenuti il miglior regressore è Random Forest sul dataset clustered\_plus\_prolog\_generated.csv

## Capitolo 10 : Conclusioni e Sviluppi Futuri

In questo progetto, il framework costruito ha dimostrato l'efficacia nell'analisi delle prestazioni di accelerazione, integrando algoritmi di machine learning .Le operazioni di preprocessing hanno permesso di selezionare solo le variabili più significative , migliorando così l'accuratezza delle predizioni dei regressori. I modelli hanno mostrato l'importanza di alcune variabili come la potenza del motore e la velocità massima, fondamentali per il tempo di accelerazione

Come sviluppi futuri, si potrebbe pensare di ampliare il dataset, magari con dei parametri ambientali che possono condizionare il tempo di accelerazione