

Report Regressione Lineare

I. Introduzione

L'oggetto del seguente report è quello di valutare l'opportunità di un investimento nel settore "Media" negli Stati Uniti d'America. Per fare ciò sono state prese in considerazione ed analizzate le relazioni che intercorrono tra la variabile target "EPS", ovvero l'utile per azione, e un insieme di variabili di tipo finanziarie associate ad una lista di aziende negli Stati Uniti. Attraverso una regressione lineare è stato possibile ottenere una stima dell'utile a 12 mesi per azione in base alle caratteristiche delle aziende.

II. Data Preprocessing

La prima operazione è stata quella di verificare il numero di aziende incluse in ciascuna "Sub-Industry", con un'attenzione particolare al settore "Media". Da questa analisi sono risultate solo 45 aziende appartenenti a questo settore. Dato il campione ridotto, si è deciso di includere nell'analisi anche sub-industrie simili al settore oggetto di studio.

Effettuando un'analisi di cluster ⁽¹⁾, sono state incluse nel campione anche le sub-industrie "Industrial Support Services", "Personal Goods", "Software and Computer Services", "Telecommunications Equipment" e "Industrial Transportation".

Tutte le analisi successive sono state effettuate su un dataset con 441 osservazioni al netto di outliers ⁽²⁾.

III. Regressione Lineare

In via preliminare, si è proseguito alla standardizzazione del dataset per armonizzare le diverse scale di misura delle variabili e risolvere eventuali problemi di ordine di grandezza. La variabile dipendente è stata tuttavia mantenuta con i suoi valori originali per preservarne l'interpretazione.

La regressione lineare ⁽³⁾⁽⁴⁾⁽⁵⁾⁽⁶⁾ottenuta riporta la seguente espressione:

$$\text{EPS} = 3.08 + 2.76 \times \text{EBITDA} + 0.32 \times \text{RETURN_ON_ASSET} - 1.53 \times \text{ENTERPRISE_VALUE} + 0.31 \times \text{RETURN_ON_INVESTED_CAPITAL}$$

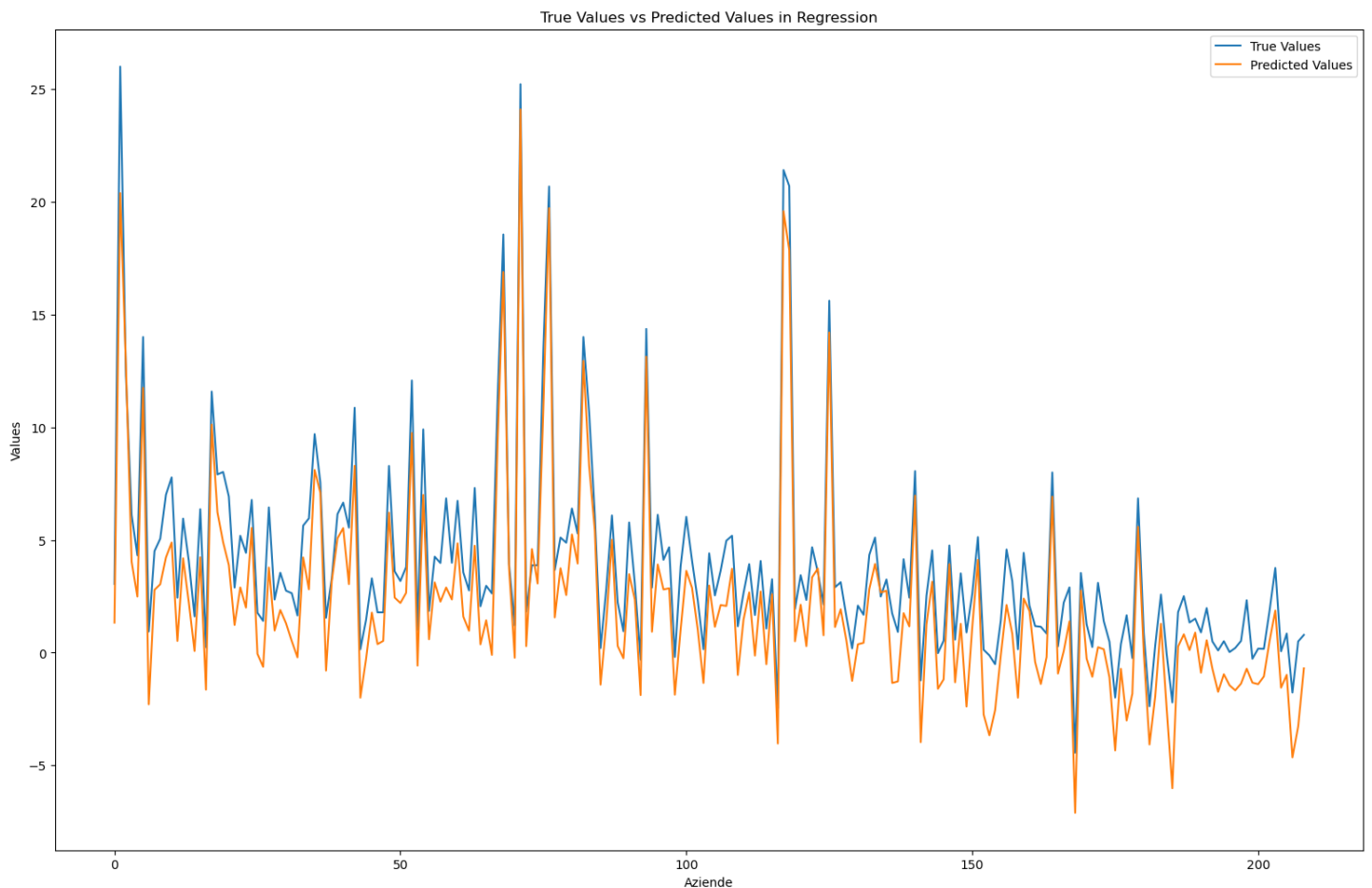
EBITDA: L'aumento di un punto EBITDA può essere tradotto con l'aumento di 2.76\$ di EPS. Essendo l'EBITDA una misura della redditività aziendale che non tiene conto di interessi, tasse, ammortamenti e accantonamenti è logico affermare che un suo aumento sia positivamente correlato con una crescita della redditività dell'azienda.

Return on Asset: Il coefficiente risulta positivo per un valore di 0.32, ciò può essere giustificabile dal fatto che un aumento di risorse disponibili per gli azionisti vada ad impattare positivamente per un valore di 0.32\$ sul guadagno per azione.

Enterprise Value (EV): Per questa variabile risulta un coefficiente di -1.53. Questo potrebbe essere spiegato da diverse motivazioni: un'azienda con un EV elevato potrebbe avere minori prospettive di crescita futura rispetto a un'azienda con un EV inferiore. Ciò potrebbe portare ad un minore EPS a lungo termine. Questo potrebbe anche essere dovuto ad una valutazione sotto la pari del prezzo delle azioni dell'azienda da parte del mercato.

Return on Invested Capital (ROIC): Questo indice consente di misurare la redditività del capitale investito nell'azienda. Un incremento di un unità di questo indicatore ha un impatto positivo sull'EPS, che accresce di 0,31\$. Un maggior ROIC porta ad una maggiore redditività e quindi ad un aumento degli utili per azione.

Per rendere più intuitiva la bontà delle predizioni, è stato inserito il seguente grafico che mette a confronto i valori reali della variabile target EPS con quelli predetti ottenuti grazie al modello. Si nota come la predizione sia soddisfacente, con un MEA (errore medio in valore assoluto) di 2,56 ⁽³⁾⁽⁴⁾⁽⁵⁾⁽⁶⁾.



IV. Conclusioni

Dai risultati dell'analisi effettuata, si consiglia al cliente di investire sulla sub-industry di tipo "Media" tenendo conto dei contributi che apportano le variabili selezionate nella regressione.

Questo si traduce nell'effettuare una selezione di aziende che presentano valori maggiori alla media di ciascuna variabile con coefficiente positivo (EBITDA , RETURN ON ASSET e RETURN ON INVESTED CAPITAL) e valori minori alla media di ENTERPRISE VALUE, il quale ha un impatto negativo sull'EPS.

La seguente tabella riporta le medie per ogni feature selezionata e dunque, per poter definire l'investimento profittevole, il cliente deve avere dei valori perlomeno uguali o superiori a quelli indicati in tabella (tranne per la variabile ENTERPRISE VALUE il cui valore deve essere al di sotto della media). Qualora ciò non dovesse accadere, l'investimento in questo settore risulta sconsigliato.

Sub Industry Media	EBITDA	RETURN ON ASSET	RETURN ON INVESTED CAPITAL	ENTERPRISE VALUE
Valori Medi	0.875001	0.311429	0.781143	11.23648

Se considerassimo i valori medi di cui sopra come caratteristiche di una ideale X azienda nel settore media, utilizzando il nostro modello di regressione lineare (previa scalatura delle features), il valore stimato della variabile EPS sarebbe di ~7,88. Alla luce della media generica del valore di EPS del dataset utilizzato per la regressione (3,42) è possibile affermare che l'investimento nel settore media, tenendo conto delle affermazioni fatte in precedenza, risulta ad ogni modo essere profittevole.

V. Appendice

1. Per la scelta delle sub industry è stata effettuata una clusterizzazione di tipo K-Means con $k = 4$, che è stato ritenuto il miglior partizionamento tramite Elbow Method.
2. La rimozione degli outliers è avvenuta dopo aver verificato tramite boxplot la distribuzione di ogni singola variabile. Abbiamo impostato un valore soglia dello z-score maggiore di 2 e abbiamo quindi eliminato i records che superavano questo valore.
3. Inizialmente, è stato eseguito sul dataset un holdout (70%-30%) per addestrare il modello. Successivamente, sul dataset è stata applicata la Lasso Regression per individuare e rimuovere le variabili che causavano overfitting e comportavano collinearità. Per irrobustire i risultati del modello è stata effettuata una cross-validation sulla regressione lineare che ha tenuto conto solo di quelle variabili selezionate con la Lasso (EBITDA, RETURN ON ASSET, RETURN ON INVESTED CAPITAL, ENTERPRISE VALUE), ottenendo un miglioramento del MEA. I coefficienti della funzione di regressione indicati precedentemente sono quelli risultanti dalla cross-validation.
4. Per quanto riguarda la significatività dei coefficienti trovati, è possibile affermare con evidenza empirica che le variabili EBITDA ed Enterprise Value risultano statisticamente significative ai fini della spiegazione del modello. La variabile Return on Asset risulta essere poco significativa rispetto alle precedenti, mentre Return on Invested Capital risulta essere non significativa.
5. In seguito, sono state verificate le assunzioni della regressione lineare. E' stato effettuato il White Test che ha confermato l'omoschedasticità degli errori con i seguenti risultati: Test Statistic 22,89 e P-value di 0,06. Il Test Durbin-Watson per controllare se vi fosse correlazione tra i residui ha dato esito negativo considerando il valore prossimo a 2 (2.01). Tra tutte le assunzioni, l'unica a non essere stata rispettata è stata la normalità degli errori con un p-value dello Shapiro-Wilk Test prossimo allo zero.
6. Il VIF presenta valori nella norma, tranne per le prime due variabili EBITDA e RETURN ON ASSET che assumono valori rispettivamente di 17,28 e 13,11. Per tutte le altre variabili il

valore è al di sotto di 3.

Intervallo di confidenza coefficienti

Output Regressione	Coefficiente	Standard Error	t	0.025	0.975
const	3.08	0.21	14,66	2,74	3,59
EBITDA	2.76	0.76	6.81	2.71	5.72
RETURN ON ASSET	0.32	0.28	1.14	-0.19	0.92
RETURN ON INVESTED CAPITAL	0.31	0.77	0,40	-2.57	1.50
ENTRERPRISE VALUE	-1.53	0.28	-5.46	-1.75	0.26

Analisi a cura di:

- Alessandro Cancelliere (S1110266)
 - Alberto Sartini (S1113674)
- Leonardo Galassi (S1112174)
 - Leandro Sciuto (S1110773)