

Predictive Analytics Challenge

Hathaway Zhang xizh3366@colorado.edu
APRD 6342 Digital Advertising

I. FINAL R^2 VALUE

After training on 70% of the data and cross-validating r^2 for 5 times, I eventually got my optimal test r^2 as **0.9208814**. In this case, the final r^2 of training data is 0.9861514.

II. MODEL

In this project, I used LassoLarsCV, LarsCV, RidgeCV, ElasticNetCV, and OrthogonalMatchingPursuitCV to train my data, and **RidgeCV** with fixed alpha values turns out to be the best model. Since `random_state` will select a random feature to update, I stimulated the splitting process for 20000 times to find the best r^2 in each model. RidgeCV was able to output optimal r^2 for both training and testing data set; moreover, it provides detailed coefficients for many valuable parameters. LassoLarsCV also provides relatively good result, but in most cases, the only useful predictor is *budget*. Other models, like LarsCV, ElasticNetCV, and OrthogonalMatchingPursuitCV, didn't provide competitive r^2 in the end.

III. SUPPLEMENT PREDICTORS

After watching all the videos and identifying their features, I added 27 predictors eventually. All these predictors were manually added based on my personal review of each ad. Three of them are categorical and the rest is binary.

- *callToAction*: whether the ad includes a call to action
- *quality*: what is the overall quality of the ad [objective]
- *keywords*: does the ad displays the keywords
- *contact*: does the ad displays contact information
- *CEO*: the protagonist is the CEO or founder
- *beautifulFigure*: the protagonist is a good looking
- *celebrity*: the protagonist is celebrity or famous
- *targetedStarring*: whether the protagonist is chosen based on ad's target audience
- *animated*: whether the ad contains animation
- *brightcolorful*: whether the ads involves in colorful and bright background
- *logoInvolved*: does the logo of company/product appears
- *mimicUI*: whether the ad displays mimic UI
- *shocking*: is the story shocking and beyond expectation
- *bgmQuality*: how's the quality of bgm. 1 means no bgm, 2 means has bgm, and 3 means great bgm involves
- *eyeCatchingTarget*: contains an eyes catching target
- *conversation*: the ad is in conversation style
- *wordy*: is the narration wordy and quick
- *narratage*: the ad is in narrative style
- *textOnly*: no narration involves and only displays text
- *empathy*: does ad uses empathy to evoke situations that happen in your daily

- *humor*: is the ad humor and funny
- *introduce*: has the ad introduced the product/company
- *accent*: does the protagonist has an accent
- *suspense*: whether the ad contains a suspenseful ending
- *guarantee*: whether the ad guarantees the quality and service of product
- *diversity*: does the ad has multiple protagonists
- *category*: what is the category of the ad [objective]

IV. COEFFICIENT WEIGHTS

After iterating the model for 20000 times, I ranked the significant impact factors by counting the occurrence of predictors that have a significant coefficient. Eventually, I find the top 20 factors by descending weights are: ***budget, placement cost, Charity and Philanthropy, targetingCost, DIY, Real Estate, 000-\$374, doctor, Undergraduate, reach, narratage, beautifulFigure, sport, school, awareness, Online Activity, Autos, Company Sales, SEO, and Pets & Animals***. For the specific model I chose, the top 5 critical predictors are:

predictor	coeff.
school	82.796
Undergraduate	81.743
Business	66.602
Business Education	61.874
Company Sales	-56.396

An interesting fact is LassoLarsCV model generates exactly the same top 20 factors as RidgeCV model does.

V. MANAGERIAL TAKEAWAYS

A. Factor That Drives Ad Clicks

Charity and Philanthropy, school, Undergraduate, Business, Business Education, etc — these factors are all coming from the attribute *audienceInterest*. According to the coefficient weights I computed, it is clear that, besides budget and cost, the audience's interest is an essential factor that might motivate the audience to click the ad. Moreover, enhancing the quality of the ad, especially regarding narration and characters, might also increasing the clicks.

B. Tailoring Current Ad Campaigns

Personally, since *Charity and Philanthropy* is a significant topic that audiences are interested and our non-profit client is a senior care organization, I think our current ad campaign should put more keywords relevant to philanthropy to raise the awareness of our client. Text ad and extension should also include words like "charity" and "volunteer". Those figures that might occur in our ads should look credible and friendly to attract more targeted audience.