Phillip Roman

Task 4: Predicting a Facebook Post’s Number of Users

Facebook is the largest social media product in the world. Business’ are utilizing social media to not only market but service their diverse customer base. Reaching a greater number of users increases the platform’s use of service engagement. After a digital marketing post is published, Facebook logs key metrics that can be used to gage the success of the campaign. This analysis focuses on key components of a post and attempts to predict the number of users that 1) the post reaches, 2) consumes the post, and 3) engages with the post.

The input features are post category, page total like, type, post month, post hour, post weekday, and if the post was paid. There are three types of user metrics that were predicted. Lifetime Post Total Reach (LPTR) is the number of unique people who saw a page’s post. Lifetime Engaged Users (LEU) is the number of unique people who clicked anywhere in a post and other interactions. Lifetime Post Consumers (LPC) is the number of people who clicked anywhere in a post.

I began the analysis by tuning three model types – Random Forests, Support Vector Machines, and Linear Regression. When comparing r2 scores for both base and tuned models, the highest performance 0.25 for Lifetime Post Consumers was using a tuned SVR.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Metric** | **Model** | **Tuned** | **r2** | **Feature Selection** | **Additional Features** |
| LPTR | LM | N | 0.03 | N | N |
| LEU | LM | N | 0.17 | N | N |
| LPC | SVR | Y | 0.25 | N | N |

An r2 value of 0.25 is not indictive of a good model fit. Only 25% of LPC’s value is explained by the model’s inputted features. This is where finding for an Exploratory Data Analysis come into effect. There is a positive relationship between impressions and Lifetime Post Consumers. Based on the sample data set, there are 3.62 post consumers for every 1,000 impressions. Impressions are not interactions and are not calculated on as an interaction with a post. Also, an RFE recommended removing the paid and category features.

|  |  |
| --- | --- |
| Feature Importance | |
| 1) Lifetime\_Post\_Total\_Impressions | 0.50 |
| 2) Page\_total\_likes | 0.13 |
| 3) Type | 0.11 |
| 4) Post\_Hour | 0.10 |
| 5) Post\_Weekday | 0.08 |
| 6) Post\_Month | 0.05 |
| 7) Paid | 0.02 |
| 8) Category | 0.01 |

After repeating the tuned and untuned models, r2 increases to 0.46 by adding impressions as a feature and .70 when also removing the paid and category features.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Metric** | **Model** | **Tuned** | **r2** | **Feature Selection** | **Additional Features** |
| LPC | SVR | Y | 0.25 | N | N |
| LPC | SVR | Y | 0.46 | N | Y - Impressions |
| LPC | SVR | Y | 0.70 | Y | Y - Impressions |

Although an r2 of 0.70 indicates a large proportion of Lifetime Post Consumers is explained by the reduced number of inputted features, the fact that r2 was only able to be increased above 0.50 by utilizing the impressions metric is not good for future predictions. Although the r2 is high for the model, the model is not useful in production. Lifetime Post Total Impressions is only known after a page has posted content. The analysis does reiterate the concept that engagement is driven by impressions. Businesses should invest in advertisement which increase impressions if their goal is to increase the number of post consumers.