Name - Sayem Lincoln

PID – A54207835

SSC 442 Exam 1

Workforce Analytics

1)What might be a good application of statistical learning in this environment? What is the

object we are predicting, and what are the inputs?

Answer – Statistical learning is a good way used to determine associations between predictor(independent) and response(dependent) variables, the statistical learning in this environment would be- y=Beta0+Beta1(x1) + u. Here y is the dependent variable (response) and x variables are independent(predictor), Beta0 ( intercept ) refer to a value of y when x=0 , while Beta1 (regression coefficient , also we call it the slope) refer to the change in variable y when the variable x change one unit, and u the error term.

The object we are predicting here are the number of candidates who got hired and how they are progressing after being hired – whether they were a success or got fired. So, the inputs would be – number of people who got hired, number of hirees who succeeded, and who did not.

2)We need to be able to justify our decision-making. Describe how would we interpret

the output of your answer to #1 when speaking with senior management. If the methodology you chose in #1 above is not easily interpreted, explain why.

Answer – There are 2 type of outputs that can be delivered from the regression model. The first one being the succession coefficient, and the second one being the failure coefficient. By doing this Human Resources can figure out the rate at which they should employ and will also present them with a predictive model by which they can figure out how many of their hirees will be a success and how many will be a failure. Additionally, if Beta1(succession) > Beta1(failure) then they would know that the hiring process resulted in a workforce benefit, and if it was Beta1(succession) < Beta1(failure) then it led to a workforce loss.

The succession coefficient can be outputted by regressing the number of hired employees against the number of hirees who succeeded – by doing this we get a coefficient value (Beta 1) which is the success coefficient. For the failure coefficient, the predictor variable has to be changed from number of hirees who succeeded to number of hirees who failed, and a coefficient Beta1 value for failure would be delivered.

Echo Team (Engineering):

1. We’re looking to bring new engineers in, but we want to make sure they can do some simple

programming. Can you think of an example of a simple visualization that they might do using our data? How would this deliver business value to the unit? Your answer must include reference to the relevant R package for the visualization.

Answer – Create a bar chart on the number of errors that you received, categorize the errors by type, then present the categorized errors on the x axis and the number of times you received the error on y axis. For example, the categorized errors can be – Search problems with…

1. Audio Search,
2. Video Search,
3. Voice commands,
4. Geographically wrong output,

Etc…

By plotting a graph of these variables against the number of times Alexa faced problems doing these activities, the programming department can prioritize which troubleshooting they need to deal with, so that they can better user experience.

For plotting a graph like this, and in this instance a bar chart, we can use the ggplot2 library from R, additionally the versatility of ggplot2 will enable us to add additional features to the existing graph too, we can add the rate at which these problems had occurred by adding a scatter plot above the bar chart that demonstrates the rate of the error, and further features can also be added due to the library’s versatility.

Twitch Data Science (Analytics):

1. What is a simple linear model we could run to look at which characteristics of streamers

encourage donations? What are some inputs X and what might our target variable Y be?

Answer – Simple linear model would be y=Beta0+Beta1(x1) + u. Here y is the dependent variable (response) and x variables are independent(predictor), Beta0 ( intercept ) refer to a value of y when x=0 , while Beta1 (regression coefficient , also we call it the slope) refer to the change in variable y when the variable x change one unit, and u the error term.

In this case, y= number of bits received, x= content type - content that influences more members of the public. Additionally, diverse content would reach more viewers and would attract more viewers to watch their content increasing the streamer’s ability to earn more bits. So, as their content influences more people then they will receive more bits, the rate at which they will receive bits is determined by the coefficient Beta1, the greater the coefficient the more bits the streamer will earn.

2. We are thinking we might conduct a small-scale experiment in which we change the price on

donations. How large of a sample (in terms of the number of users) do you think we need? How

does this depend on your answer to #1?

Answer – If the population size was 20,000 subscribers then the sample size should be about 400, with a confidence level of 95% as the confidence level is the amount of uncertainty we can tolerate and a 5% error margin level – as the margin of error gives us the amount of error that we can tolerate, now if the margin of error decreases then surely the sample size would have to be increased as they are inversely proportional to each other. Finally, the response distribution should be taken as 50% - as creating an unbiased model will give us the best model for the streamer’s content.

So, taking the model from #1, and running the model within the above-mentioned statistical parameters a regressed model can be achieved that delivers an outcome with both prediction and confidence bands. Without being run on a regression model as mentioned in #1, a regressed outcome would not be possible as it would not present Twitch with a predictive model on how the streamer might do in real life on the platform. If the streamer is already doing well then their viewers will not be opposed to the increase in bits price as they surely like the streamer and their content, but if the streamer is on a downfall (losing viewers or getting unfavorable feedback from the viewers on the content that they are publishing) then increasing the bits price can surely impact the streamer in a negative way, as then the viewers will be opposed to buying more bits and sending it over to the streamer.