Written Task – Responsible Gambling

Responsible Gambling

Sports wagering is an enjoyable socially but can be harmful of not monitored. I appreciate BetTube’s efforts to come up with an algorithm to detect harmful gambling behaviors among their regular customers and help them keep gambling in check. This would also keep the customers loyal and help the company gain their trust as it shows that the company is not trying to extract huge amounts of money from the customer.

The most important metrics that can be helpful in detecting a gambling problem in customers are the behavioral metrics of the customer. Since there are transactions happening every time a person gambles, a lot of useful data can be collected to monitor whether the person is susceptible to develop or already has a gambling problem.

Metrics currently captured

As I am not completely aware of sports wagering and the key factors used in wagering, I’m not quite sure I understand all the numeric metrics given in the raw csv. However, assuming that each observation or row in the file indicates factors involved in the gamle - AFL (Hitouts, Home Team/Away Team, Kicks, handballs etc) and not the details of the customer himself, not a lot of variables can be used to develop a responsible gambling algorithm. Having said that, since the data is captured for seven years, one key metric that can be utilized from this data is the frequency of gambling for each customer. We can create another database which would be customer centric, the primary key (unique value) would be the customer id which can be used to identify each customer. Then, for each customer, the total number of times the customer has gambled, the frequency of gambling of the customer in a year, month, week, day can be calculated. This calculated frequency would tell us a lot about the gambling behavior of each customer. These metrics can then be used to build a classification model to classify the customers as being a potential gambler or not.

Frequency of gambling of customers can be used to calculate various other factors like – daily frequency, weekly frequency etc. Maximum gambling sessions on a daily, weekly and monthly basis. This would indicate how often the customer is gambling. If the daily or weekly frequency of the customer is high, or the maximum number of daily sessions in a customer is consecutively high, it would clearly indicate a problem. The frequency of the customers can also be used to calculate the gap between gambling sessions. Metrics like average gap between session, shorted gap between sessions, longest gap between sessions could be some important metrics to look at to predict gambling problem in customers. The duration or the time spent by a customer in each gambling session can also indicate an issue. If the customer is frequently gambling on all days including weekdays and not just weekends or weekdays can also be used.

Metrics than can be captured

Apart from the metrics currently being captured (columns in the raw csv file), there are lot of other metrics that can be captured to predict if a customer has or is developing a gambling problem. The other main thing that can be captured is the money a customer spends on gambling. Customers with a problem might spend a lot on gambling, frequently. Therefore, their expenditure amount and rate are important metrics. For example, if a customer spends a big amount in every gambling session, the maximum and amount he spends in a gambling session. A customer’s behavior after he loses a big amount in wagering is a key indicator of a problem. The urge to win back all the money lost in a session makes a customer gamble more with high stakes. This compulsion to win back the money despite losing repeatedly is usually the start of a gambling problem. If in case the customer does win back a huge amount after losing it first, the adrenaline rush makes the customer repeat this behavior. Therefore, the customer repeatedly gambles with high stakes despite a loss streak. The number of losses of a customer can be correlated with the expenditure amount. If there is a correlation between the two it might mean that the customer wants to win back the money after every loss and therefore tends to gamble again.

Predicting if a customer is susceptible to a gambling problem is a classification problem. A suitable machine learning classification problem can be built using the metrics stated in the above paragraph. There are various machine learning algorithms for classification – Logistic Regression, Decision Trees, Linear Discriminant Analysis, Neural Networks etc. As it is important to understand the importance of each variable on the probability of the customer having a gambling problem, we need to use algorithms that can indicate the significance and magnitude of effect on the response variable. Here, the objective is to understand the relationship between the independent variables and response variable while also trying to accurately predict the probability of the response variable. Therefore, stochastic models like Logistic Regression and Linear Discriminant Analysis etc, would be a better fit as we can easily obtain the magnitude of coefficient of the independent variable. On the other hand, non stochastic processes like Decision Trees or Neural Networks would treat the model as a black box with minimum or no indication of the relationship of each independent variable with the response variable.

After identifying the customers with a potential gambling problem, limits for spending money and time in each session can be determined. Once the model (to detect gambling problem in cutomers) is finalized and deployed, warnings can be issued to the customer if he gambled above a threshold or is gambling for a long time. By triggering these limits, we can make sure that the customers are aware of their gambling traits and limit their gambling and enjoy it socially