Дима, привет. Смотри что нужно сделать, вот ссылка на код

Идея работы такая

Каждый вечер модель нужно переобучать, т.к. копится стата по операторам, на худой конец раз в неделю .

Обучащаюся часть кода 1-370 строку, если смотерть в notepad++

Т.е. берутся данные из базы и поехали.

Строка 11-18 Важно помнить, что в файле orders-answer.csv нет названия полей, в коде они прописываются согласно порядку, который Смыков сказал. В исходных данных, откуда бы вы их не брали из базы напрямую или через выгрузку csv, необходимо чтобы поля стояли таким образом.

dt <- fread("orders-answer.csv",

header = FALSE,

col.names = c("lead\_create","lead\_id","order\_id","approved\_at",

"called\_at","webmaster\_identifier","product",

"operator\_login","operator\_gender","operator\_age",

"operator\_country","lead\_country","answer",

"postsale","parent\_order\_id"),

na.strings = c("", "\\N"))

Я не знаю какая там у тебя база, но обычно подключаются вот так.

library("RODBC")

dbHandle <- odbcDriverConnect("driver={SQL Server};server=ТВОЙСЕРВЕР;database=ТВОЯБАЗА;trusted\_connection=true")

#запрос

sql <- paste0 ("Select то что нужно from нужная таблица”)

df2 <- sqlQuery(dbHandle, sql)

итак после обучения модели генерится файлмодели

349 строка # Сохранение и загрузка модели

save(glmnet\_model, file = "output/glmnet\_model.RData")

Все готово

Теперь модель начинает работать на свеже поступивших юзерах

Что она должна сделать

Пришел лид с товаром

Теперь в строки 378-385

Нужно системе нужно проставлять автоматически в R

1. В поле product вставить название товара пусть это для примера будет ChocolateSlim
2. В поле n\_calls\_before вставить число звонков , которое этому лиду делалось. Т.е. если лид не взял трубку с первого раза то в след раз для этого же лида в поле n\_calls\_before будет стоять 1 и так далее. Т.е. для каждого лида обученая модель может в день выдывать разные вероятность в зависимости от числа звонков.
3. В поле hour\_create указывает промежуток времени когда пришел лид, если он пришел в промежутки 14-15 часов то в поле hour\_create ставится начальное число промежутка ,в этом примере оно 14 часов.
4. time\_diff это предполагаемое время звонка, т.е когда оператор собириется ему звонить. Если это 20 минут, то ставим долю от часа, т.е. 20 минут это 0,33 часа, ставим это число и логарифмируем
5. hour\_call это час звонка. Если ему позвонили с 14-15, то час стоял бы 14, в моем примере ему позвонили в 15:05)) час звонока 15.

test\_data <- data.table(product = "ChocolateSlim",

n\_calls\_before = 0,

time\_diff = log(0.33),

hour\_create = 14,

hour\_call = 15)

Далее

Идем в строку 588

И в product вставляем название товара, которое лид заказал

newdata <- dt[product == "ChocolateSlim", ][1, ]

в итоге генерится фалй result.csv там вероятности для этого лида

Все с этим лидом закончили и такая же участь ждет каждого лида))

Обрати внимание что все что идет со строк 375 по самый конец запускаются для каждого лида отдельно. А также для каждого лида отдельно если изменились параметры

n\_calls\_before = 0,

time\_diff = log(0.33),

hour\_create = 14,

hour\_call = 15