Outreach KT

<https://j9zovp.axshare.com/#id=6diete&p=member-outreach>

<https://onlinect.sharepoint.com/:f:/s/StarRatingPOC-SDE2.0MemberOutreach/Ehcyqc8ApFRPrKv9NWNctkEBs1vtXGelwyUxxDmYppIcCw?e=BX1x4m>

<https://10.30.5.81/ShrutiV/sde-member-outreach/tree/develop>

Decision engine:  
is nothing but a product designed by CTS Tech to help payers. Pairs are nothing but insurance companies, and like every year, government gives some star rating to each payer they are like it is compulsory to publish them in newspapers. So that accordingly, according to the star rating, people can choose their insurance for the upcoming year and also if the star rating of a particular insurance company is improved even by .5.  
Start, then they get a huge amount of money from their government for their improved performance. So each payer wants to improve their performance and there are different majors on which the performance is calculated every year

Center for Disease Control has published a list of measures that are utilized to calculate the performance every year. Different measures are used. It is not fixed, which all majors will be used for calculating the performance of a pair in this particular year. It is dynamic, they keep changing it so.

as a insurance company or as a payer they want to improve their performance because in they will be benefited in two different ways.

One they will get money from the government and other is they will get more members. So eventually their revenue will increase.

So they start Decision engine is a product which helps them identify which are the measures or factors on which they should work on. It identifies top factors and then it gives these are the factors on which if you will work on.

perform plus   
 member outreach

member outreach was a supervised model. It was a classification model

three classification models.

OK, so as I was saying on SDE 1.0, there is a white paper. It is like OK, so under HP innovation garage.

There is a channel called like SD 2.0 member outreach

SD 1.0 data and code base

Who bring up HD 2.0 version where along with what SD 1.0

these are the five factors on which if a payer works, they will improve their style rating say 5.5 say by one star and so on. But then what are the factors on which they like? How should they work on it? Is what like this member outreach POC is supposed to help with.

Outreach

what is outreach is so suppose we are payers, we are an insurance company and we have a number of Members who have taken the insurance from us. Now it is responsibility or like one of the important.  
Ohh revenue gaining and performance gaining part of for **payer is get outreaching all the members** and then **prioritizing their health check up** and then getting like their aim is to have maximum of their members to be healthy.

This is like this is their aim so so they want to outreach the people.

so that they will be more compliant towards their what are the whatever prescribed?

A health checkups are so here we want to find out what is the **best channel to outreach the Members.** What is the **best time to outreach** them so that the **outreach will turn out to be positive** and then the **Member would turn out to be actually compliant**. But since we are dealing with a pair, they have like like thousands millions and billions of members who would be associated with them. So then **it is not really possible to outreach every one of them.**

Effectively**, that is why first so we so this member outreach program is made-up of three levels of models**.

**First one will figure out** it will give **a group of people who have higher probability of responding to the outrage**. So so for that we have built one model and then using these models.

So these people who what is **the best time to outreach them** and what is the **best method to outreach** them is what this particular POC.

**Member like member, prioritization will be done** more like **risk stratification, clinical segmentation, uh disease, category segmentation**.

All of this **is taken care by SD, 1.2 or SDE.** **That product which Citiustech has**, but then member **outreach is the one where we are working**. It will **understand how Members are engaged across different touch points** based. Like I explained, **communication channel and time.**

0:11:37.640 --> 0:12:9.720  
so for a particular member we identified, OK, this person has higher probability of responding to the outreach and their outreach channel and time has been identified. But what to do after outreaching them? What will be the?

Like for example, uh, I found out that **there is a person whose age is about 60** and they have like **diabetes and hypertension** and some other disease as well. Then I'm outreaching them by call and like in the morning, this is what my model gave. But then what should I like outreach them for? What should I talk about like go go for these tests or what? What are their limitations because of which they are not compliant. So this is something which.

It was still in the scope, but we did not work on it, so that that's the next part. We will go on. So what is in the **scope of this particular POC is best identifying best communication channel and best, best time to outreach content** was not actually included in and the assumptions for were target member data will be already available.

A target will be **member outreach for the POC** and **not provider outreach,** so this will be the Members and not the providers and **data source will be demographic, clinical and historical outreach data**. So **these members were outreached earlier and what was their response to which particular channel** did they respond which particular which particular time they're **response rate** was higher and so on kind of variables would be considered are the assumptions for this particular POC.

Then what are the outcomes? Is framework design so and the recommendations to the payers are providers then **data model creation and baseline machine learning models and reusable code and pipelines**. So these all we have created and the approach to deal with it was first research and review the existing tools and literature. How **not not only healthcare related companies but how existing member outreach is going on.**

0:14:7.910 --> 0:14:32.690  
Rutuja Kore  
Then we will create the data because it was a POC, we did not have a data available with us, so we had generated a few few **few features synthetically** and then we had followed regular **DSLC steps** like data **quality check, then EDA feature engineering, then model building validation** and so on. And then we had validated this with the internal stakeholders.

Data we had utilized, but obviously it did not have historic data. So we had used the data from **Microsoft.** we did not have a few variables. So those were generated synthetically.

So like **we had basically created the data the way we would want for member outreach**, so that it will be a framework which would be ready. So if you get in, if in real life you don't get that exactly same kind of data **maybe in some other forma**t, you'll have to bring it in the format this model getters or maybe model could be verified like like changed. But this is like a framework on what all variables will give like.

high level flow of this POC looks like.

first orange colored box tells us about what all kind of data we have. We have **demographic and socioeconomic data, then claims medical history.**

claims medical history data was taken from CMS deidentified data which is publicly available. Demographic data was also available. Some socioeconomic factors were not available

we had generated, for example, income was generated synthetically.

And then the data was combined then uh member risk model. So who are the high **high risk members** then? Who are **the target members**? These two are blue boxes are **done by SD 1.0.** So this is like.

So this POC is built on the **assumption that the output from a SD 1.0 will be utilized for HD 2.0 t**hat is this member outreach.

**So we had assumed that we will get the high target members** who like out of all the Members **who are the targeted members**

And

**whose compliance would turn out to be in the improvement of STAR rating**. So those Members will only be considered.

And then it was also assumed that a **health risk score and priority score from SD 1.0** would be available, but we are not really utilizing it in the model, but then it was the assumption in the beginning.

So **these are the Members who are like input format data**. So first what? So this is the older version, but later on **we had added one more channel to it like one more model** to it which out of even though **we have Member target list** out of that, **there could be Members who does like these are higher priority members and they are higher risk members** maybe or but then who all of because this list is still very big, we want to know who all.

Are the **ones which will actually respond to the outreach**.

So we **have a model which tells it like it is a classification model**. If the **person will respond** to the outreach, **yes or no**. So then that is utilized here.

As for the **best time and best channel model,** only those Members will go for it. And then the plan was best. Incentive will also be decided, which we did not work on. And **combining these three customized outreach would be done on those members.**

It was decided what we will do is **we will look at the member profile like demographic, social, economic profile, then clinical profile and in the beginning we will randomly assign some.**

Outreach response for them based on similar characteristic of people who have had some sort of response earlier because the company is supposed to have some data with them. So for the Members who are repeated who have joined their insurance, like taken their **insurance once again for them, they will have data for newer ones we won't have for them it will be assigned randomly** and then so this will be the beginning step and then whatever their response will be, it will be traced back and it will be added.

And so this loop this particular loop is supposed to continue, and then we have built a high level like a wireframe which will tell.

Ohh which will help the user interact with it and see who are the Members who have higher uh like who who will come under the **call category, morning category and so on**. We'll look at the wire frame before proceeding towards the code and all in order to understand it better but this is how the flow is supposed to go on so that our outreach tracking will also go on. So suppose if in my current outreach my model gave OK outreach.

Person in the morning with call, but he here she did not respond to it, so that should also be take like included in the learning of the model. And **next time maybe this information will help in predicting**. OK, maybe he will respond to e-mail or some other thing. So that is why this information will also be fed to the model.

So what are the benefits of this model to CT? Are like member or provider outreach framework can be used as a next version of UH D 1.0. Although the current model is for member outreach except for like we will not have member like demographic data and so on. But in some other format similar kind of information could be provided by providers. **And As for a payer, provider is also an important.**

**0:22:23.500 --> 0:22:54.100**Rutuja Kore  
Factor because Members eventually go to **providers only for their care**, so then provide a performance management is an important factor for a payer. So that's what **perform plus** in perform plus framework is about. But we will come to it later on but then same framework could be applied to providers as well. Then we can use this as outreach analytics solution like **stand alone outreach analytics solution for pairs provider departments such as care management teams.**

. Then we can integrate it with the existing platforms which were there already like BI and score plus and then market validation. So these were the identified benefits at that time.

So basically our objective was to develop an **intelligent member outreach framework** to ensure **increased member response rate** and thereby **improved CMS star rating** for health plans. So who will be The Who, who are the Members for whom the solution will predict is the **Members who are most likely to respond to the outrage and how will they respond is the best communication channel for member outreach.**

are the problems that are there and how are we dealing with dealing with it using this framework so lack of prediction while communicating with members so we all get calls from say Flipkart, Amazon or different chains that OK we have this offer and go for it. It is different from healthcare outreach but still we get and many a times we tend to ignore the emails or we don't even pick up the call.

0:24:27.810 --> 0:24:59.380  
Rutuja Kore  
Even if we pick up, we don't actually go and shop. So since our target members are 60 plus audience like 60 plus members, there could be like younger population as well. But it is, it was observed that the maximum revenue is from the 60 plus age group people, they could be the ones who who might not be like physically enabled to go and get a call. They might not have access to smartphones. So they might not be reading the emails.

0:24:59.470 --> 0:25:3.460  
Rutuja Kore  
And that is why it is not like converting into.

0:25:4.510 --> 0:25:35.140  
Rutuja Kore  
Positive response or entirely anything like this was like this is exist which will help pair to track all these things then missing analytical angle in outreach journey because they are just sending emails and they're just calling. But what is happening after that is not being tracked. So that's where this particular framework is supposed to help. What are the benefits to the customer out of it are effective outreach strategy increase in member outreach?

0:25:35.210 --> 0:25:41.750  
Rutuja Kore  
Response rate and improving star rating for health plans, which are directly associated with member outreach.

0:25:43.30 --> 0:26:11.760  
Rutuja Kore  
So then what we'll data sources like like I told earlier, demographic claims and clinical data from CMS data and then member outreach data from like we had taken it from some different kind of survey by Microsoft, it was not related to healthcare, but that's the best data we could find at that moment. So that was utilized. But in general this is the kind of data that we want and then outcome will be the recommendation various.

0:26:12.250 --> 0:26:20.840  
Rutuja Kore  
For various aspects of member outreach, like outreach responders, communication channel and time for member outreach. So this is the same flow.

0:26:21.260 --> 0:26:51.290  
Rutuja Kore  
Ohh but it is like improved version. As I said, outreach responders model is the first thing that we start with, then best channel time and then customized outreach, then outreach responses stored and it is fed back then outreach tracking dashboard which tells attention, interest, desire and action related things. So we will have a look at the dashboard again but before that we will try to understand on a high level how do.

0:26:51.490 --> 0:27:24.0  
Rutuja Kore  
These three models look like and what were their outcomes? So first one was the model which predicts the target responders. So our aim was to build a machine learning model to predict if a Member will respond to the outreach campaign. Here we are not considering what kind of mode, what time and so on. It is just on whether they will respond yes or no. So input parameters where demographic sdoh, clinical data and historic outrage data. So the model which we have trained is based on 34,000 members.

0:27:24.420 --> 0:27:55.300  
Rutuja Kore  
And the output was yes or no and so the best model like best algorithm which was giving the like 93% accuracy, 73 percent F1 score 83% recall and 66% precision. So these attributes were given by random forest. So final model was random forest for the given data and the value delivered by this particular model was prioritizing on members who are more likely to respond to an outreach.

0:27:56.10 --> 0:27:58.390  
Rutuja Kore  
So the this is how the important.

0:27:59.60 --> 0:28:29.470  
Rutuja Kore  
A features look like, which contributed in the response. So first one was first three are historic responses call, e-mail, SMS. So we have considered three channels of outreach call, e-mail and SMS and then we have considered three times of outreach for like morning, afternoon and evening. So those were also fed to this model. So we can see that education.

0:28:29.640 --> 0:28:59.640  
Rutuja Kore  
Number of dependents they have household size, gender. Then again, historic attempts and like so you can see there are two things I don't know if you can see clearly historic response. So if the response was yes or no and how many attempts were given. So for example for call I had called a particular member 10 times out of that I got response only three times. So response is recorded by this particular variable historical response and the.

0:28:59.790 --> 0:29:26.20  
Rutuja Kore  
Number of attempts for particular channel have been recorded by historical attempts and then the name of the channel and then total hospitals in that the area where they are residing, age, median income state in which they're residing marital status. So these were the top features contributing to the decision on whether the Member will respond yes or no. So like on a high level, this is how the.

0:29:27.280 --> 0:29:32.50  
Rutuja Kore  
Model look like we will go into the details of data and code and so on but.

0:29:32.590 --> 0:29:39.340  
Rutuja Kore  
Uh, like very high level this. This is how it looks like. Is it OK? Do you have any question?

0:29:44.940 --> 0:29:45.240  
Rutuja Kore  
No.

0:29:40.310 --> 0:29:47.200  
Uma Maheswari K  
Rutuja actually responses and one more thing is attempts right, but attempts should be more than responses, right?

0:29:47.730 --> 0:29:48.260  
Rutuja Kore  
Yes.

0:29:49.460 --> 0:29:53.290  
Uma Maheswari K  
But uh, here in the graph which it is showing less.

0:29:54.10 --> 0:29:57.710  
Rutuja Kore  
No, no. So it is like it is, this is importance plot.

0:29:58.920 --> 0:29:59.340  
Uma Maheswari K  
OK.

0:29:58.380 --> 0:30:2.170  
Rutuja Kore  
It is not the count plot. It is like so.

0:30:8.200 --> 0:30:10.350  
Uma Maheswari K  
OK. OK. Yeah, yeah, I got it. Yeah.

0:30:15.80 --> 0:30:15.870  
Uma Maheswari K  
Yeah, yeah, yeah.

0:30:10.430 --> 0:30:21.120  
Rutuja Kore  
So response plays more important role than attempt. I may be calling 10 times, but I'm not responding so it will go down. So this graph basically means that.

0:30:21.470 --> 0:30:23.160  
Uma Maheswari K  
Peter importance like.

0:30:22.810 --> 0:30:33.520  
Rutuja Kore  
Yeah, yeah, yeah. Correct, yes. And then these colors respond on whether the, uh, like, in which category this particular variable favors more.

0:30:34.770 --> 0:30:43.950  
Rutuja Kore  
But here we can see almost equal distribution for both responders and nonresponders, but that's added information. This graph is supposed to give.

0:30:45.180 --> 0:30:45.650  
Uma Maheswari K  
OK.

0:30:45.380 --> 0:31:12.460  
Rutuja Kore  
OK. So moving on, we will see how the best predict like prediction of best channel model works on so again on the similar lines, we had similar kind of data like 34 K Members demographic, ADHD clinical and so on. The output words cold calling e-mail and SMS we had considered only these three channel considering they're easy to track. And then the second thing is.

0:31:12.540 --> 0:31:42.710  
Rutuja Kore  
Uh, we considering the age group of our targeted population. We had considered these three channels then that technique was like the best model which we found out with like 89% accuracy, 87 percent F1 score and 86% recall and 89% precision was extra tree classifier. And then the value delivered out of this model was identification of key attributes in predicting best channel to outreach. So it will.

0:31:43.130 --> 0:31:53.390  
Rutuja Kore  
Give you what is the best channel for a given member, but we it will also help you identify what are the factors which are contributing in a particular type of.

0:31:54.360 --> 0:32:26.10  
Rutuja Kore  
Like channel being the best channel for that member. So like along with doing classification, if you go and look at the like this particular graph you will you can also do kind of clustering in the background like these are the group of people who respond to cold calling more and what are their characteristics. So that like I said for new coming members we can assign them in the first we have to assign them something so there.

0:32:26.100 --> 0:32:40.150  
Rutuja Kore  
It is supposed to help, so here you can see again response variables are coming in top then number of children how many children they have. This is because our targeted population was 60 plus.

0:32:41.230 --> 0:33:10.640  
Rutuja Kore  
If they have somebody to take care of them, maybe that's why a number of children is coming up, then income, obviously. Then education, number of dependents, age, then depression indicators like if they will respond to a given channel or not. Maybe like their this was coming up annually income. What is the best time to outreach them disability indicator and marital status and other like CHF like chronic heart failure.

0:33:10.730 --> 0:33:33.300  
Rutuja Kore  
Indicator if they have severe heart disease. So basically comorbidities if there are so like we had different variables related to comorbidities as well, but these are the ones which are coming up in top. So historical outreach response I like I said these are the variables which are coming out to be the top contributors and here you can see the difference for example.

0:33:34.770 --> 0:33:50.720  
Rutuja Kore  
Historical response rate call is coming out to be top, but it has more impact on cold calling type of outcome which makes sense. Then again for mail also you can see pink color has a higher like broader.

0:33:51.710 --> 0:33:59.680  
Rutuja Kore  
Area covered in this bar, which means like which is obvious like it is compliant with which we will otherwise expect.

0:34:0.740 --> 0:34:29.950  
Rutuja Kore  
Uh, similarly, we had another model for uh, best prediction time. So the goal was to build a model which will predict best time for member outreach to ensure increased member response rate. Same kind of variables were considered here as well. But we had six outputs like weekday, morning, weekday, afternoon, weekday, evening and then weekend morning, weekend, afternoon, weekend, evening. So we had six.

0:34:31.510 --> 0:34:41.210  
Rutuja Kore  
Categories which will be predicted by this model. Here Random Forest was giving the best outcome for our.

0:34:41.510 --> 0:35:11.240  
Rutuja Kore  
Hmm. Majors, performance majors, and the value delivered by this model is the identification of key attributes and predicting best time to reach out for member engagement. Here again, the response rates for six categories is coming out to be in top for like these classes are nothing but weekday morning, weekday, afternoon and so on, which are in the same order like weekday morning, weekday afternoon, weekday, evening 012 and then weekend morning weekend.

0:35:11.360 --> 0:35:13.820  
Rutuja Kore  
Kanan weekend, evening, SO456.

0:35:14.700 --> 0:35:19.890  
Rutuja Kore  
Ohh again similar sort of variables are coming out to be in top like earlier two models.

0:35:21.320 --> 0:35:30.100  
Rutuja Kore  
Uh, so this is a screenshot of how the wireframe looks like. We will instead go and look at the wireframe itself.

0:35:31.70 --> 0:35:42.880  
Rutuja Kore  
And so it it shows like it will have different campaign drop down. I I don't think it is. OK. So here we had only one for example I want to work on cancer awareness.

0:35:43.60 --> 0:36:15.730  
Rutuja Kore  
Get campaign and then what are the majors? So here only colorectal cancer screening has been considered. So for this the Member was member, we want to outreach members for this campaign and particularly this measure. This will be the output of SDE 1.0. Then how many total members who were outreached then their response propensity. So this is basically given by my first model high, medium low. We are utilizing it in terms of yes or no but then it.

0:36:15.820 --> 0:36:46.340  
Rutuja Kore  
There is also giving us like basically we are getting the probability of response. So yes or no is defined according to the threshold that will go towards when we will look at the code but then we have also converted it into categories high, medium, low. So eight people out of 780 are the ones who are giving high propensity. So this is a dummy kind of data that we had fed. But this is how it will give the results then how many of people responded to calls emails.

0:36:46.410 --> 0:36:52.910  
Rutuja Kore  
SMS then timings, weekday, weekend this this all is in percentages and then.

0:36:54.370 --> 0:37:25.800  
Rutuja Kore  
We can choose who are the I can choose only high higher response members, then low, medium and so on. And mode of communication can also be filtered. Time can also be filtered and similarly the list will be modified. So all like like the outputs AI suggestions for response. That's how we are tracking here response propensity mode of communication and time. So that's how it is showing so. But in order to look at the member profile we have another.

0:37:25.900 --> 0:37:58.140  
Rutuja Kore  
A A like frame here where I can filter on member propensity mode of communication and time. Same as earlier and it gives me high level idea on what are the characteristics of the people residing in that particular group like age, gender, marital status so that like I said earlier we can sort of in background identify the properties of the members who respond to particular kind of outreach so that this will help in defining the care journey for that particular person.

0:37:58.360 --> 0:38:11.610  
Rutuja Kore  
Then what is the distribution of chronic conditions, depression, diabetes, heart disease, kidney disease and so on? Then what are the different types of disability types that these people have? So it is like if I.

0:38:12.370 --> 0:38:42.740  
Rutuja Kore  
If I filter on, uh, high propensity people and mode of communication as call and time as uh morning weekday so it is not changing here but it is supposed to change then we will get an idea on what is the disability type that these people have. What are the social determinants like, household size, education, income and so on. It is supposed to tell us. Then we have model explain ability dashboard. So let's have a look at best propensity.

0:38:42.930 --> 0:39:1.380  
Rutuja Kore  
So OK, you will have to run this. I think I will have to have that file with me. So what this particular. Sorry, what this particular third model explain ability dashboard does it we have integrated the output of model explainable.

0:39:1.450 --> 0:39:3.690  
Rutuja Kore  
Pretty, uh are.

0:39:3.790 --> 0:39:22.80  
Rutuja Kore  
Ohh library which gives like interactive graphs related to what one we are getting here, but the better ones like all other features as well. Particular tools that gender and all so we can do that we'll have a look at it when we will look at the.

0:39:23.170 --> 0:39:47.990  
Rutuja Kore  
Uh code and the data. But uh, like this is how these three models are supposed to open up here. So that's how on high level, this member outreach is all about. I will give you the access to this particular channel so that you can go through the files in the documents in proposal PPT. These files are there and I'll share this part with you.

0:39:50.120 --> 0:40:2.480  
Rutuja Kore  
So next time maybe we can go and have a look in detail on how we built the models, what, what, how did the data look like? What are the codes and so on so.

0:40:3.640 --> 0:40:4.690  
Rutuja Kore  
Any questions?