- **The platform is freely available but is not fully utilized, this is based on insights acquired from Forgood platform data**

49% of Needs did not get any response

50% of registered volunteers did not take any volunteering action

Only 2% of the volunteers accessed the platform in 2019

Only 24% of causes accessed the platform in 2019

- **The wining causes has the following characteristics**

They rate their volunteers

They have high average ratings

They are situated in Gauteng

They post need that requires time with background

They are in the categories of health, ‘women, children and youth’, Christian religion, drug addiction recovery ,and law advocacy and politics

- **Volunteers clustering for segmented communication**

9578 No action

3644 Occasional

5854 Consistent

- **Platform activities forecasting**

There will be an increase in activities in 2020

activities are likely to occur on Mondays, Tuesdays & Fridays,

Busy Months

February, July, November, December

Profiling volunteering transactions in South Africa.

INTRO

- The paper discuss the use of machine learning and the Forgood volunteering data to develop tools that will improve the volunteering transaction success in the platform .

QUESTIONS

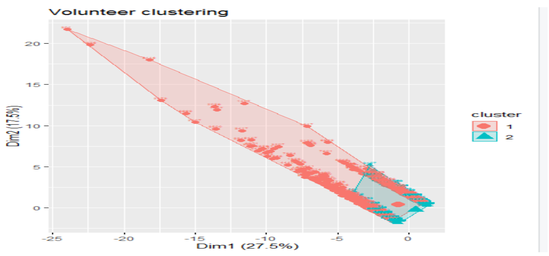
- What are the winning characteristics of a wining cause?
- Who are the occasional and consistence users?
- Can machine learning techniques be used to predict the need success rate?
- Can machine learning techniques be used to predict the volunteer's status?

METHODS

1. Logistic regression(LR) model is used to identify the characteristics of a winning causes and estimate the probability of a need getting a response
2. K means clustering method is used to group volunteers based on their volunteering behaviours.
3. The random forest model predicts the volunteers segmentation and engagement to causes to identify those that are not actively participating in the volunteering platform.

RESULTS

- Clustering of volunteers based on their volunteering behaviour, cluster 1 and 2 represent consistent and occasional volunteers.



	Cluster	average_rating	total_actions	status_active	status_deactivated	status_unverified
1	1	3.6423436	2.274149	0.4223381	0.0024698134	0.5751921
2	2	0.2725658	1.233174	0.3189272	0.0006832935	0.6803895

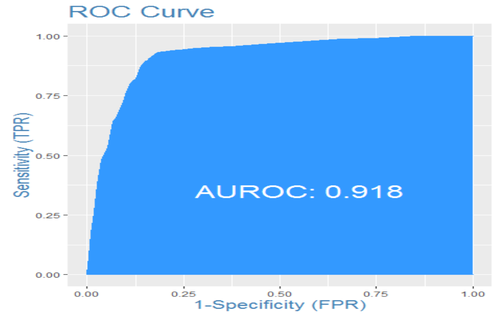
- Significant test with LR to identify the winning characteristics of a winning causes.

```
call:
NULL

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9206  -0.4056   0.1720   0.5317   3.1729

Coefficients:
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                             -4.624e+01  3.901e+02  -0.119  0.90565
average_cause_rating                     1.116e+06  1.665e+05   6.703 2.05e-11 ***
cause_quality_engagement_score          -1.256e+05  1.553e+04  -8.086 6.16e-16 ***
de_activated_needs                      -2.017e+05  1.447e+04 -13.940  < 2e-16 ***
active_needs                             5.132e+04  2.918e+04   1.759  0.07864 .
expired_needs                           -2.423e+05  2.352e+04 -10.305  < 2e-16 ***
completed_needs                          1.831e+05  6.014e+04   3.044  0.00233 **
total_documents_uploaded                -1.386e+05  1.272e+05  -1.090  0.27593
total_team_members                      -9.031e+05  1.094e+05  -8.258  < 2e-16 ***
total_stories                            3.693e+04  3.980e+04   0.928  0.35347
direct_beneficiaries                     1.742e+00  5.751e-01   3.028  0.00246 **
percentage_of_black_beneficiaries        2.679e+04  1.337e+04   2.004  0.04508 *
percentage_of_black_women_beneficiaries -3.226e+04  1.729e+04  -1.866  0.06199 .
totaluser_rating                         8.881e+04  5.401e+03  16.443  < 2e-16 ***
cause_status_x_approved                  1.073e+08  1.168e+09   0.092  0.92681
cause_status_x_deactivated               1.043e+08  1.168e+09   0.089  0.92888
cause_status_x_declined                  8.913e+07  1.168e+09   0.076  0.93917
website_no                              -8.790e+05  9.872e+05  -0.890  0.37324
facebook_no                             -1.655e+06  1.125e+06  -1.471  0.14127
twitter_no                              -1.809e+06  6.619e+05  -2.734  0.00626 **
logo_uploaded_no                        -3.007e+06  2.217e+06  -1.356  0.17509
last_dashboard_access_recent             1.050e+06  2.192e+06   0.479  0.63173
last_dashboard_access_not_recent         5.282e+06  1.803e+06   2.930  0.00339 **
need_background_no                      -2.276e+06  8.933e+05  -2.548  0.01084 *
type_time                                4.913e+07  6.399e+05  76.771  < 2e-16 ***
province_x_gauteng                       1.045e+07  1.351e+06   7.736 1.02e-14 ***
province_x_eastern_cape                 -2.127e+06  1.852e+06  -1.148  0.25086
province_x_western_cape                  2.514e+06  1.407e+06   1.787  0.07390 .
cause_category_women_children_youth      7.008e+06  1.472e+06   4.762 1.92e-06 ***
cause_category_women                    -3.601e+07  9.014e+06  -3.995 6.46e-05 ***
cause_category_drug_addiction_recovery   1.984e+07  7.571e+06   2.620  0.00879 **
cause_category_economic_development      5.119e+06  4.468e+06   1.146  0.25193
cause_category_religion_christian        2.069e+07  2.965e+06   6.978 3.00e-12 ***
cause_category_law_advocacy_and_politics 2.265e+07  3.328e+06   6.806 1.00e-11 ***
cause_category_health                    1.037e+07  1.087e+06   9.540  < 2e-16 ***
cause_category_sports                    2.941e+06  2.363e+06   1.244  0.21340
cause_category_education_research        3.061e+06  8.794e+05   3.481  0.00050 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 27276  on 19678  degrees of freedom
Residual deviance: 13327  on 19642  degrees of freedom
AIC: 13401
```

DISCUSSION

- The data points in both cluster are well structured with the average silhouette coefficient of 0.2 and 0.6 for cluster 1 and 2 respectively.
- The logistic regression model was able to accurately predict 87% of the testing data with 90% area under the curve.
- The outcome of the model provides a classification status of each volunteer, segmenting between active and unverified with accuracy score showing model predicted outcome is right 73% of the time.
- The model also provides features that have an impact in identifying volunteers segmentation.
- The stability of the models were tested using the stability function in R, at different learning samples the model is stable.

- Receiving Operating Characteristic(ROC)



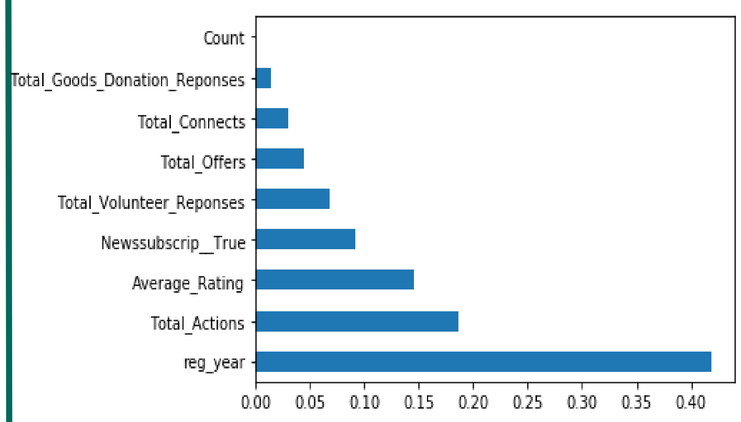
- Confusion matrix

```
Confusion Matrix and Statistics


pred    0    1
   0 1550  153
   1  331 1902

               Accuracy : 0.877
                 95% CI : (0.8664, 0.8871)
    No Information Rate : 0.5221
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7526

 Mcnemar's Test P-Value : 8.593e-16

            Sensitivity : 0.8240
            Specificity : 0.9255
         Pos Pred Value : 0.9102
         Neg Pred Value : 0.8518
             Prevalence : 0.4779
         Detection Rate : 0.3938
   Detection Prevalence : 0.4327
      Balanced Accuracy : 0.8748

       'Positive' Class : 0
```

- Silhouette coefficient

	cluster	size	ave.sil.width
1	1	3675	0.20
2	2	5823	0.61

- Feature importance



Future work include building a model that will automatically match the volunteers to the needs available in the platform, based on the outcome the platform will notify them.

Minah Manyama , Maria Twala

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA