Quality of Care

for Chronic Disease Management

VAIBHAV GAIKWAD // 2018HT12597

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE
PILANI (RAJASTHAN)

April 2020

Nobody ever figures out what life is all about, and it doesn't matter. Explore the world. Nearly everything is really interesting if you go into it deeply enough.

- RICHARD P. FEYNMAN

Introduction

- In healthcare domain the applications are build around the idea of providing care the patients.
- Philips VitalHealth has many chronic disease management solutions build for various diseases (i.e. Diabetes, COPD, Asthma etc.).
- Philips culture supports innovation to add more value to the products for staying ahead of the competition.
- One of the widely talked Unique Selling Point (USP) is about Quality of Care (QoC) in the solutions.
- The dissertation work focuses on the research of building a model for the assessment of QoC.

Motivations

Analysis of Quality of Care (QoC) is about gaining information about the effectiveness of care management for the patients.

- QoC is stated good when the health of population is stable or improving.
- QoC also is a measure of the effectiveness of a treatment protocol.
- QoC in a business case means proving the ROI (return on investments) to customers.
- QoC measurement is must for accountable care as per CDC's Meaningful Use policy.

Observations & Objectives

Observations

- Current solutions are having little or no way to assess the QoC.
- Knowledge related to QoC is observed in the people but not implemented.
- Ad-hoc methods are adopted for QoC report generation on need basis.
- Expected data is present in the existing system.

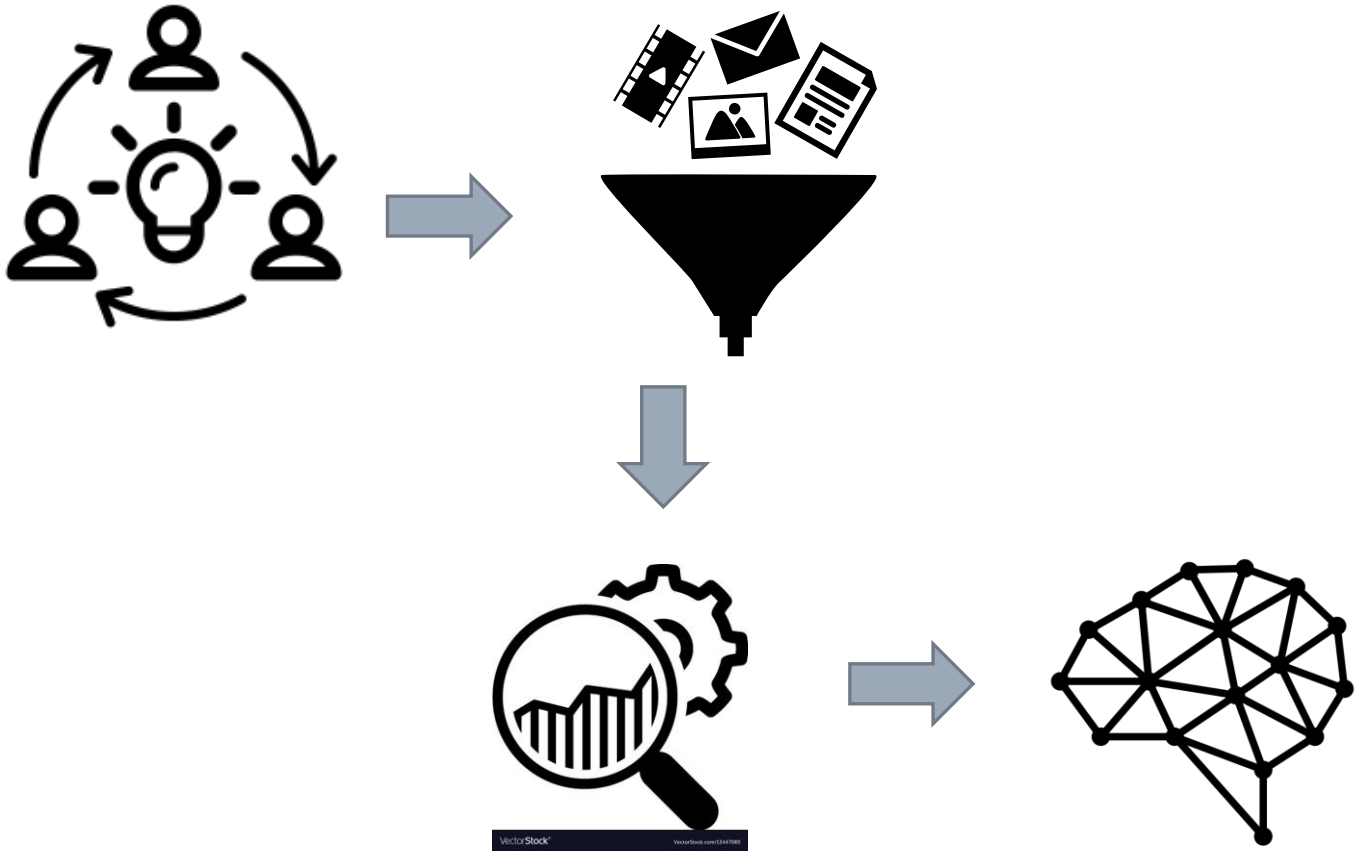
Objectives

- Define a way to find QoC attributes for an application
- Propose a model to assess the QoC using those attributes

Work Summary

Division of work

- Ideation and design
- Data collection and analysis
- Searching for risk groups
- Design for Quality of Care



Ideation

Brainstorming sessions to understand the ways to find QoC.

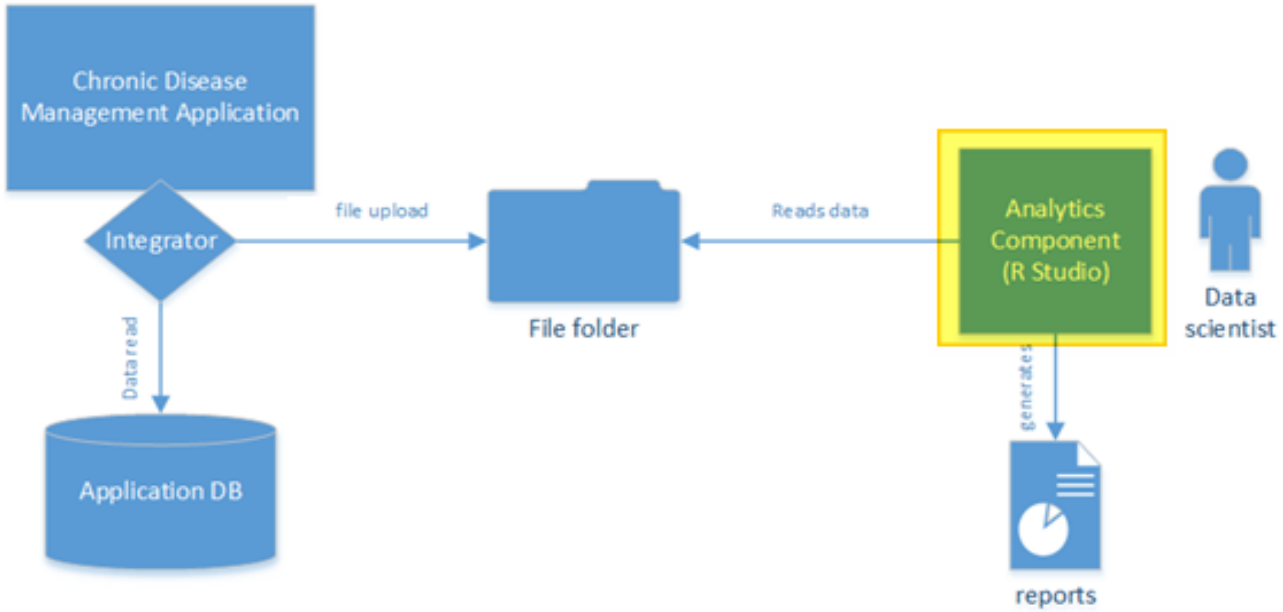
- Summarizing the patient feedback at the overall level gives information about the quality of care.
  - The adoption of such a method does not seem to be proactive.
- Trend analysis of each patient's health can also provide details on the Quality of Care.
  - This method is time-consuming and does not generate insights on the population level.
- Compare data from previously used applications which have proved better Quality of Care.
  - It is difficult to achieve due to the dependency on a trustable existing system.

Research requirement was gathered and a proposal was made with the objectives.

- Research a way to find QoC attributes for an application.
- Propose a model to assess the QoC using those attributes.

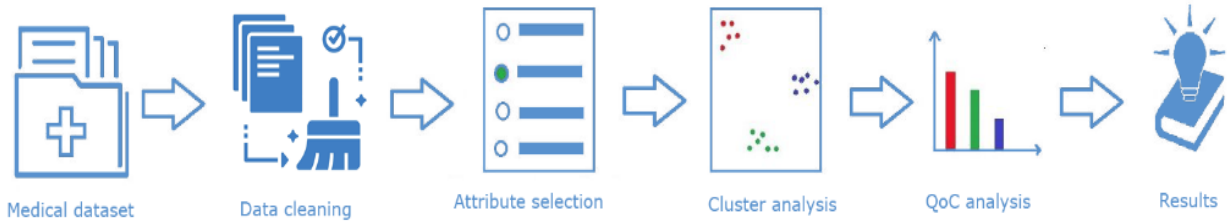
Ideation

System block diagram



Ideation

Analytics Component



- CDM holds the medical dataset, export to CSV is done.
- Data cleaning removed the insignificant data from the exported dataset.
- Attribute selection helps to find the valuable data attributes helpful for clustering.
- Cluster analysis helps to understand the risk-based groups formation.
- QoC analysis will help in understanding if the care management is effective or not.
- Results will be the reports that are generated for the care organisation.

Data Analysis

Collection

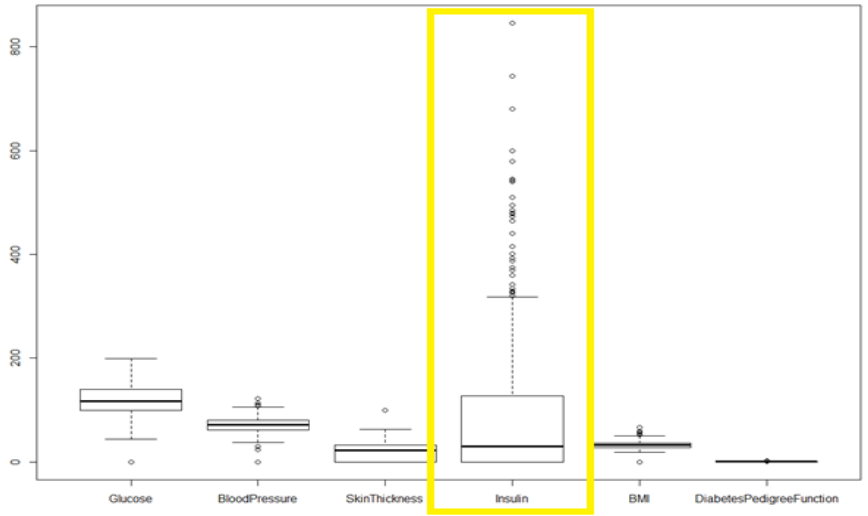
- Privacy contracts made it difficult to use the applications from live data.
- Dataset from Kaggle.com was used referred as PIMA dataset, which had 786 records for diabetic and non-diabetic female patients.
- Following eight data attributes were present as listed in the table.

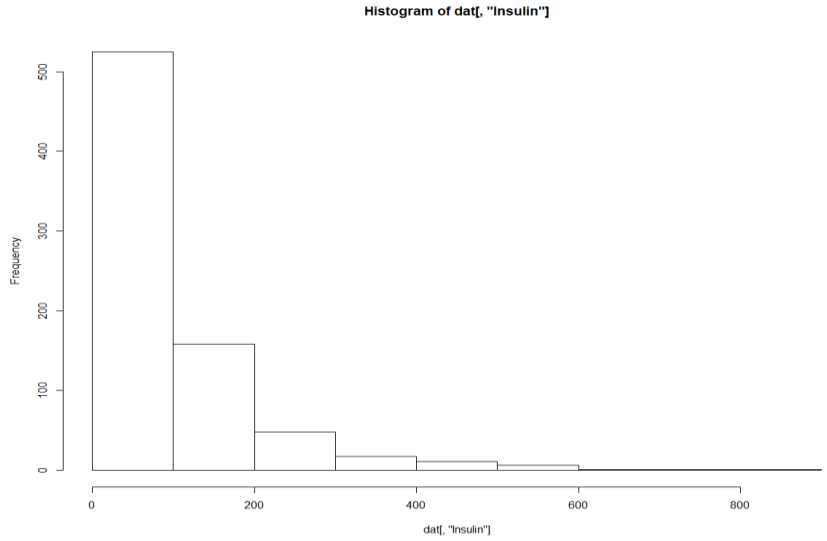
Data attribute	Description
Pregnancies	Number of times pregnant
Glucose	Plasma glucose concentration at 2 hours in an oral glucose tolerance test
BloodPressure	Diastolic blood pressure (mm Hg)
SkinThickness	Triceps skinfold thickness (mm)
Insulin	2-hour serum insulin (mu U/ml)
DiabetesPedigreeFunction	Function value which scores the likelihood of diabetes based on family history
BMI	Body mass index (weight in kg / (height in m) ^2)
Age	Age for a person in years
Outcome	Class variable (0 or 1) whether or not a patient has diabetes

Data Analysis

Cleaning

- As study was not gender specific “Pregnancies” value was neglected.
- Removal of records with 0 value for Glucose, Blood Sugar, BMI.
- Number of were reduced to 249 after the data cleaning process.





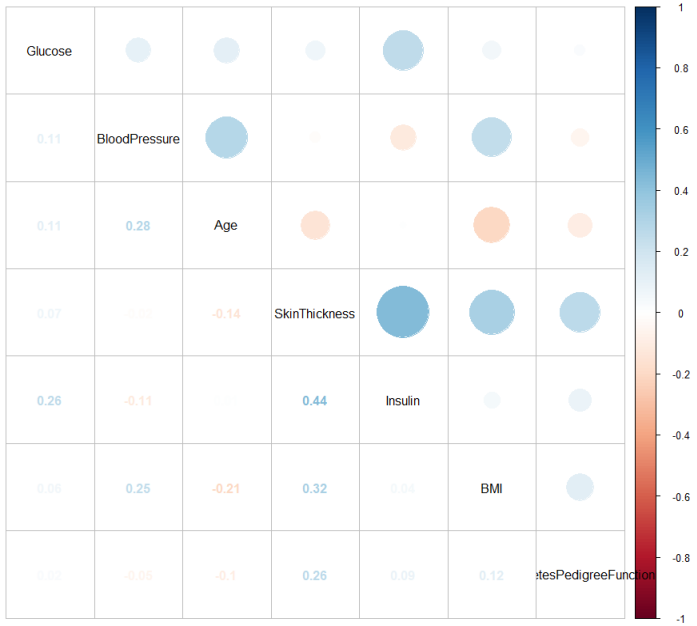
Search Risk Groups

Correlations among attributes

Correlations in filtered data with strong correlations are listed below

- *Glucose* and *Insulin* (obvious)
- *SkinThickness* and *Insulin*
- *SkinThickness* and *BMI*

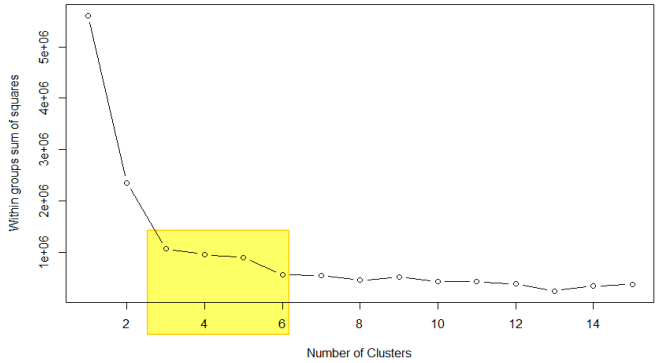
	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
Glucose	1	0.152589587	0.057327891	0.33135711	0.221071069	0.1373373	0.26351432
BloodPressure	0.152589587	1	0.207370538	0.088933378	0.281805289	0.041264948	0.239527946
SkinThickness	0.057327891	0.207370538	1	0.43678257	0.392573204	0.183927573	-0.113970262
Insulin	0.33135711	0.088933378	0.43678257	1	0.197859056	0.185070929	-0.042162955
BMI	0.221071069	0.281805289	0.392573204	0.197859056	1	0.140646953	0.03624187
DiabetesPedigreeFunction	0.1373373	0.041264948	0.183927573	0.185070929	0.140646953	1	0.033561312
Age	0.26351432	0.239527946	-0.113970262	-0.042162955	0.03624187	0.033561312	1



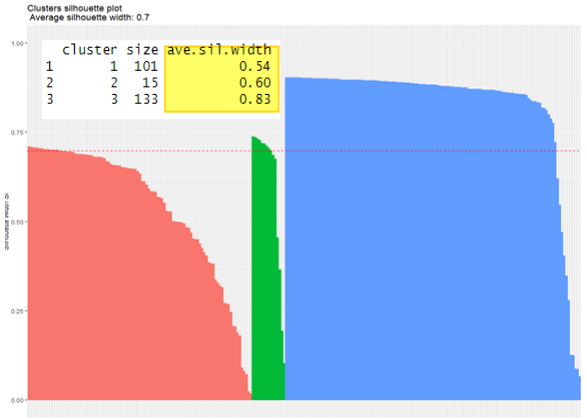
Search Risk Groups

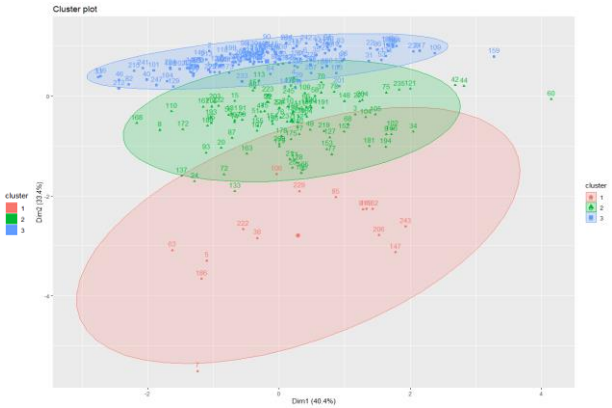
K-means clustering & Silhouette analysis

- Elbow analysis confirmed K = 3 lies in optimal section.
- Silhouette analysis iterations were carried for K [2-6] to find the attributes which resulted in good clustering in K-means.
- Attributes [Insulin, BMI, Age] were found promising at value 0.7 for Average Silhouette Width (ASW).



								Filtered data	Raw data
Glucose	Insulin	BMI	Age	SkinThick	BloodPressure	D P Func	K	Silhouette	Silhouette
	x	x	x				2	0.61	0.66
	x	x	x			x	2	0.6	0.63
	x	x	x		x		2	0.59	0.64
	x	x	x	x			2	0.56	0.6
x	x	x	x				2	0.61	0.66
	x	x	x				3	0.7	0.65
	x	x	x			x	3	0.68	0.6
	x	x	x		x		3	0.67	0.62
	x	x	x	x			3	0.61	0.56
x	x	x	x				3	0.7	0.65
	x	x	x				4	0.7	0.63
	x	x	x				5	0.66	0.62
	x	x	x				6	0.62	0.44





Search Risk Groups

Range formation

The Insulin based range formation was visible.

- Cluster 3 has a low range [0-9].
- Cluster 1 has the moderate range [96 - 328].
- Cluster 2 has the high range [360 - 846].

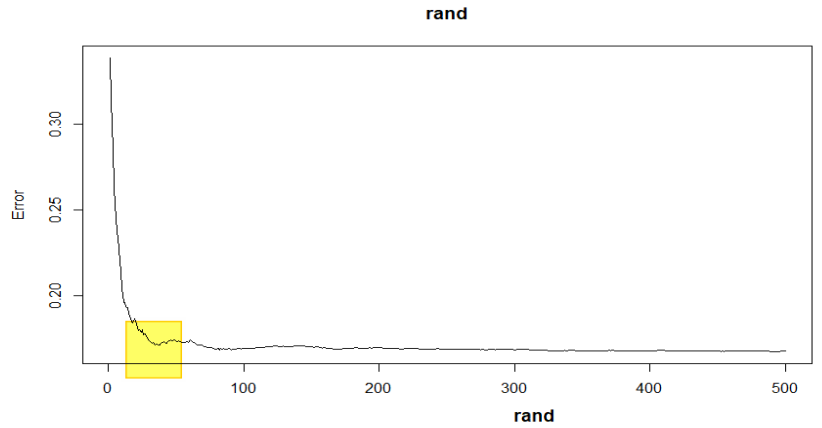
The result was not promising so attribute selection with Random Forest method was performed and clustering was redone.

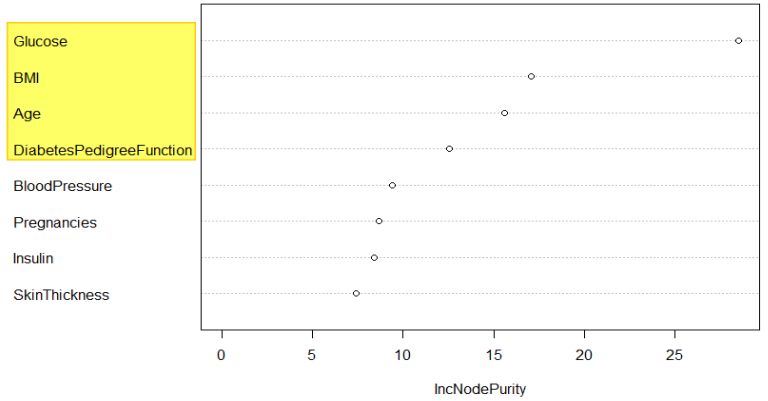
Cluster No.	Glucose	BloodPressure	Insulin	BMI	Age
Cluster 1	88 - 198	30 - 110	96 - 328	23.4 - 67.1	21 - 58
Cluster 2	124 - 197	50 - 90	360 - 846	28 - 46.2	21 - 60
Cluster 3	78 - 199	50 - 114	0 - 91	22.9 - 59.4	21 - 70

Search Risk Groups

Random Forest and Attribute Selection

- Random Forest indicated the optimal attribute combination was [Glucose, BMI, Age]. If range formation is not observed then DPF attribute can be added.
- Positive results were seen using the [Glucose, BMI, Age] after clustering.
- ASW = 0.37 was observed and was comparatively low but still in the valid range (positive).





Cluster No.	Glucose	BloodPressure	Insulin	BMI	Age
Cluster 1	78 - 125	30 - 100	0 - 258	22.9 - 55	21 – 62
Cluster 2	160 - 199	50 - 110	0 - 846	23.3 - 59.4	21 – 66
Cluster 3	123 - 159	40 - 114	0 - 600	23.8 - 67.1	21 - 70

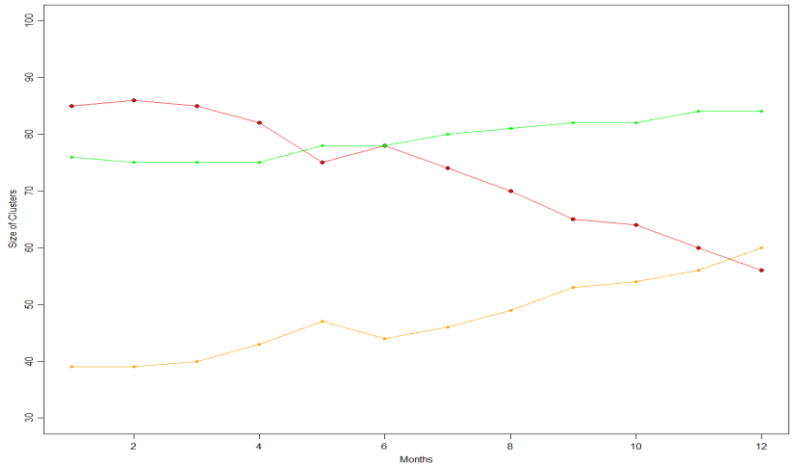
QoC design

- Clustering Confidence Score (CCS) – was defined as new terms which gives the percentage of valid data points in range to the actual world.

  CCS = [[Number of correctly identified data points in all significant data attributes in the respective ranges] / [Total number of data points]] × 10

- Design was based on trend analysis of cluster sizes that are generated from the previous steps.
- The focus is on the high-risk cluster size,
  Decreasing trend of high-risk cluster indicates good QoC.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
C1	85	86	85	82	75	78	74	70	65	64	60	56
C2	39	39	40	43	47	44	46	49	53	54	56	60
C3	76	75	75	75	78	78	80	81	82	82	84	84



Conclusion

- Time spent on improving the data quality can be reduced with better choice of datasets.
- K-means helps in analysis of risk-based groups with the right set of data attributes.
- Random Forest for attribute selection gave better results compared to ASW analysis.
- Random Forest with prediction accuracy > 80% showed risk-based groups formation.
- QoC for existing and new systems, have similar models based on trend of size of clusters.

References

- M. J. Hallett, J. J. Fan, X. G. Su, R. A. Levine and M. E. Nunn, “Random forest and variable importance rankings for correlated survival data, with applications to tooth loss,” *Statistical Modelling,* pp. 523-547, 2014.
- B. Rai, “Feature Selection and Predictive Modeling of Housing Data Using Random Forest,” *International Journal of Business and Economics Engineering,* vol. 11, no. 4, pp. 919-923, 2017.
- C. Chapman and E. M. Feit, “R for Marketing Research and Analytics,” in *R for Marketing Research and Analytics*, Springer Nature, 2015, pp. 331,332,333.

THANKS