[](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&ved=2ahUKEwjkjMDqgabiAhWLdt8KHa1NC0IQjRx6BAgBEAU&url=https://en.wikipedia.org/wiki/Saint_Mary's_University_(Halifax)&psig=AOvVaw1jOxfPyAH0z4d12foEPcn-&ust=1558301022614868)

**MSc in Computing and Data Analytics**

**MCDA 5580 – Data and Text Mining**

**Assignment – 1**

Submitted to:

**Trishla Shah**

Prepared by:

Allen Mathew - A00432526

Gyaneshwar Rao - A00433014

Meghashyam - A00432392

# Executive Summary

It should have

1. Business perspective – Idea
2. Future Steps
3. Less technical
4. Write to the point
5. What data was given – customer / products
6. Final profiling
7. Preventive measures
8. Snap shot of the work we have done

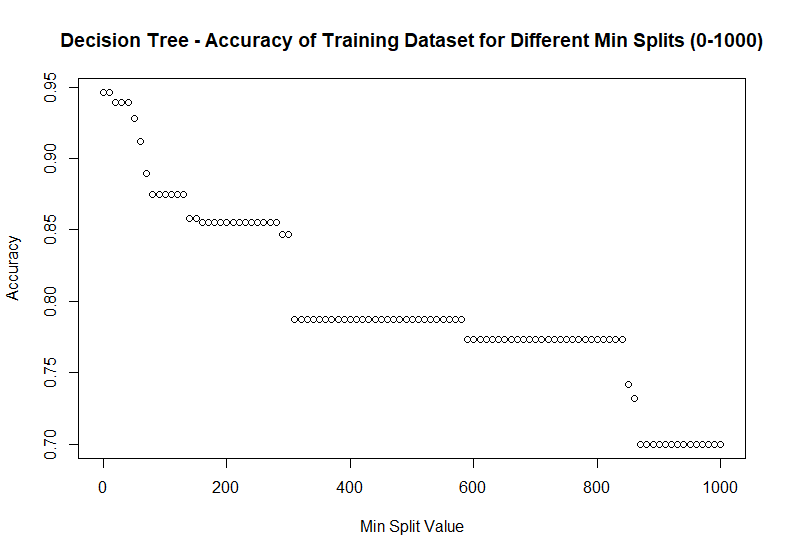
# Objective

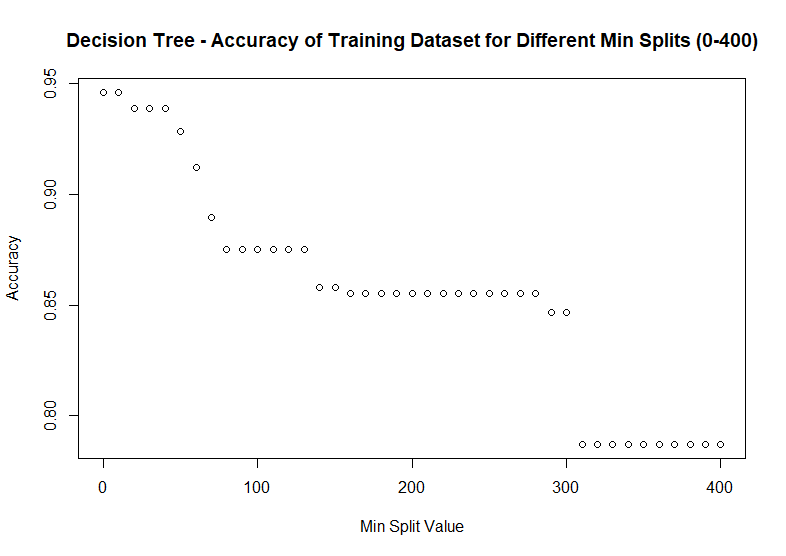
It should have

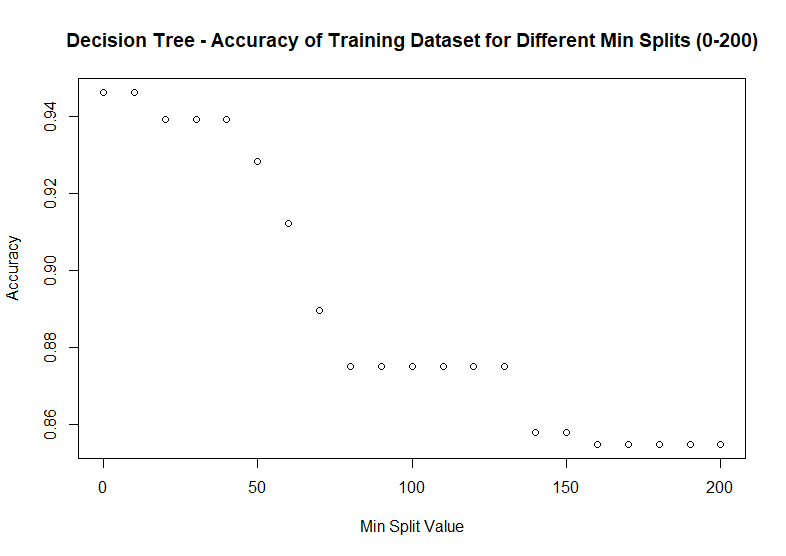
1. Technical part
2. The end point of the report
3. End goal
4. Meting the object using clustering – k means

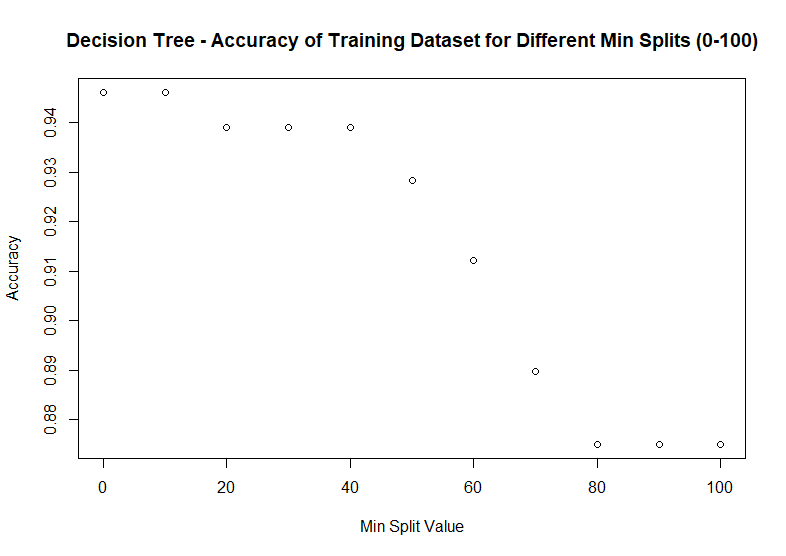
# Data Summary

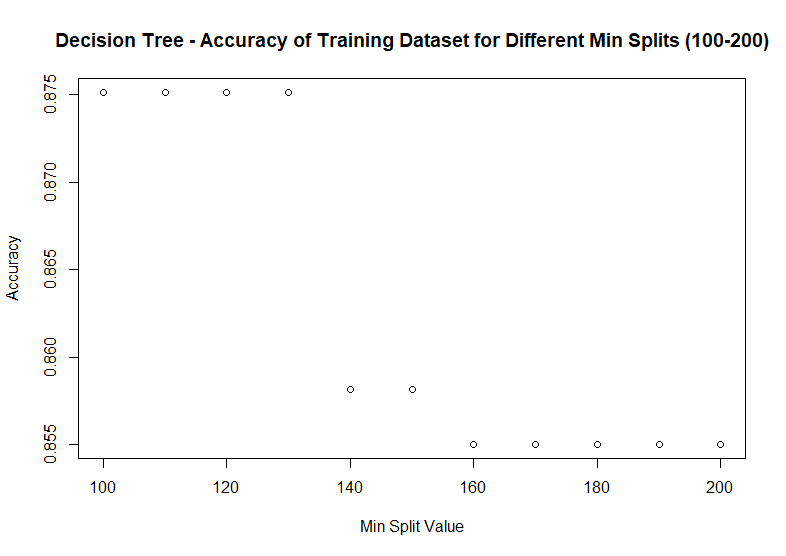
# Min Split Value (Decision Tree)

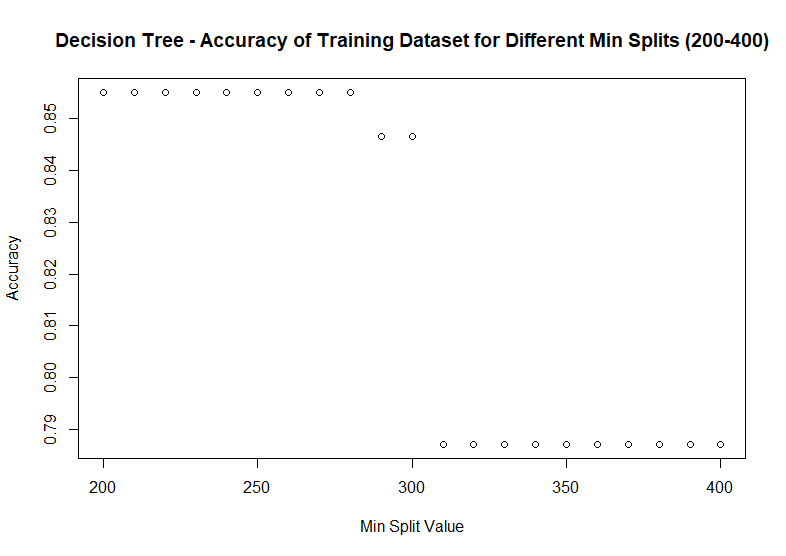


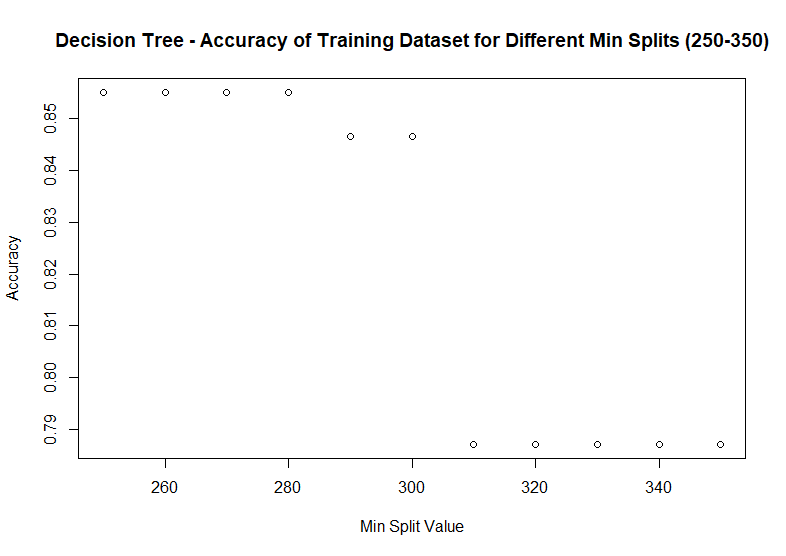












0-200

> min\_split\_list

[1] 0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200

> accuracy\_list

[1] 0.9460293 0.9460293 0.9390902 0.9390902 0.9390902 0.9282961 0.9121049 0.8897456 0.8750964 0.8750964 0.8750964

[11] 0.8750964 0.8750964 0.8750964 0.8581342 0.8581342 0.8550501 0.8550501 0.8550501 0.8550501 0.8550501

0-100

> min\_split\_list

[1] 0 10 20 30 40 50 60 70 80 90 100

> accuracy\_list

[1] 0.9460293 0.9460293 0.9390902 0.9390902 0.9390902 0.9282961 0.9121049 0.8897456 0.8750964 0.8750964 0.8750964

100-200

> min\_split\_list

[1] 100 110 120 130 140 150 160 170 180 190 200

> accuracy\_list

[1] 0.8750964 0.8750964 0.8750964 0.8750964 0.8581342 0.8581342 0.8550501 0.8550501 0.8550501 0.8550501 0.8550501

200-400

> min\_split\_list

[1] 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400

> accuracy\_list

[1] 0.8550501 0.8550501 0.8550501 0.8550501 0.8550501 0.8550501 0.8550501 0.8550501 0.8550501 0.8465690 0.8465690 0.7872012

[13] 0.7872012 0.7872012 0.7872012 0.7872012 0.7872012 0.7872012 0.7872012 0.7872012 0.7872012

250-350

> min\_split\_list

[1] 250 260 270 280 290 300 310 320 330 340 350

> accuracy\_list

[1] 0.8550501 0.8550501 0.8550501 0.8550501 0.8465690 0.8465690 0.7872012 0.7872012 0.7872012 0.7872012 0.7872012

**When taking min split 140, the accuracy of the prediction using Decision Tree is 0.8468677 or 85%**