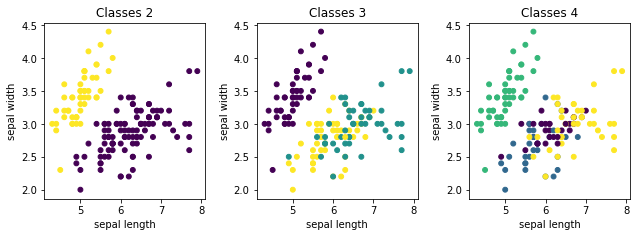
**Solutions and Screenshots**

**1.1**

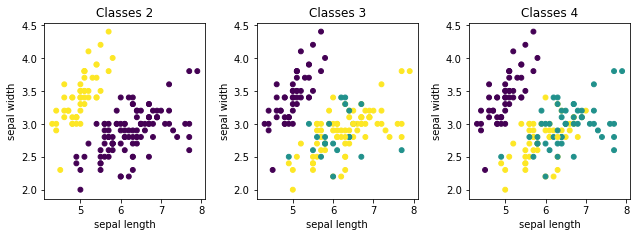
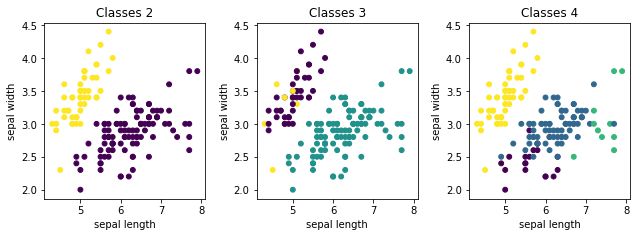
**1.2**

There is an internal criterion which states that high intra-cluster similarity is observed in good clustering. So, in case of **2** classes there is high intra-cluster similarity. This can also be done by calculating silhouette\_score. The value near to 1 indicates non-overlapping clusters while negative value states that data assign to wrong cluster. For different clusters silhouette scores are



This shows for 2 classes it is better.

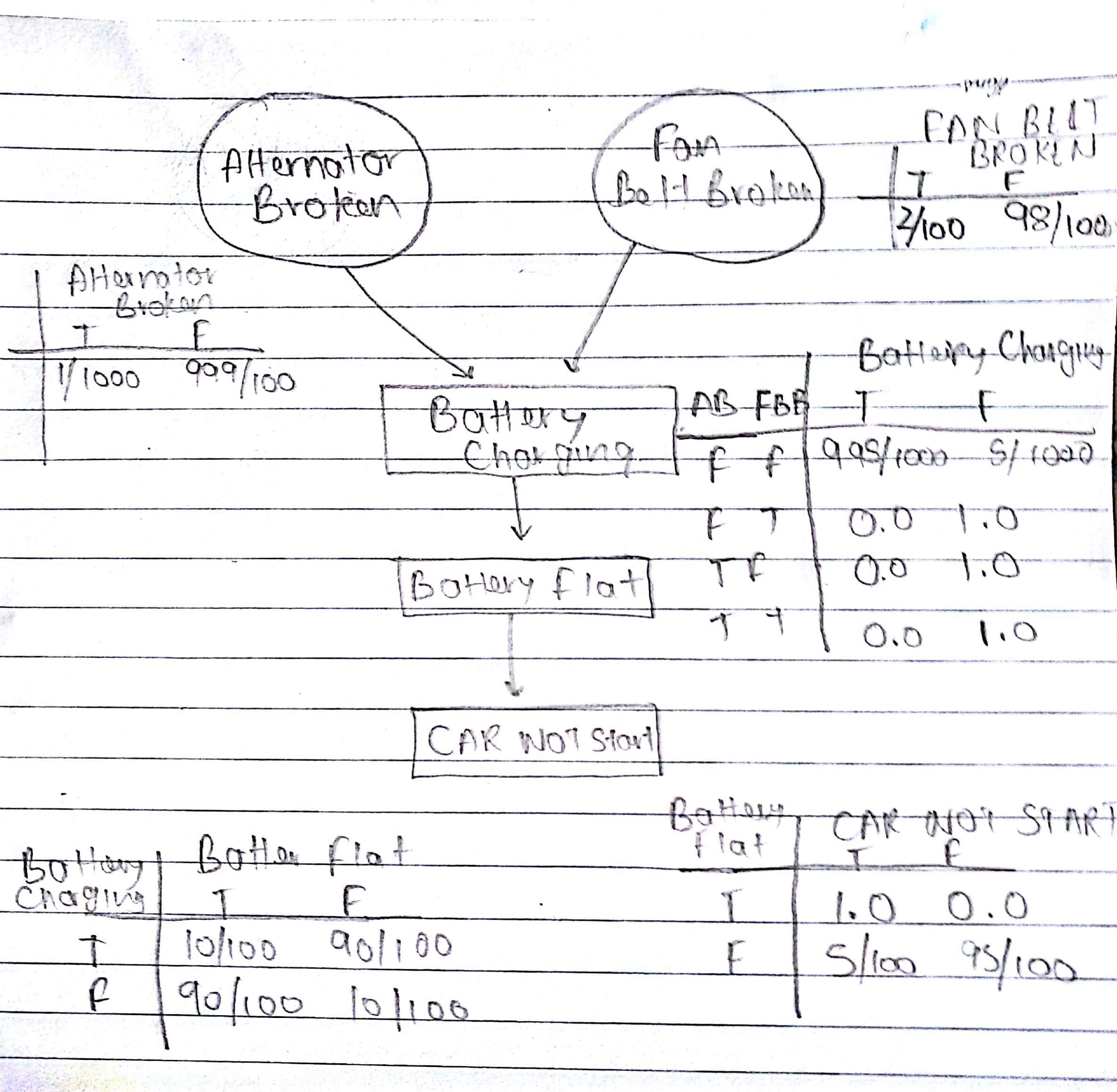
There is a stability based method [3], in which when data repeatedly produce similar clusters there is a high level of agreement. Based on this



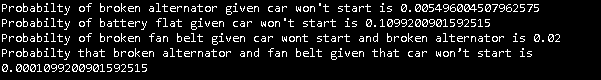
As we can see after repeated iteration clustering model (2) produce the cluster with high level of stability and agreement.

**2.**

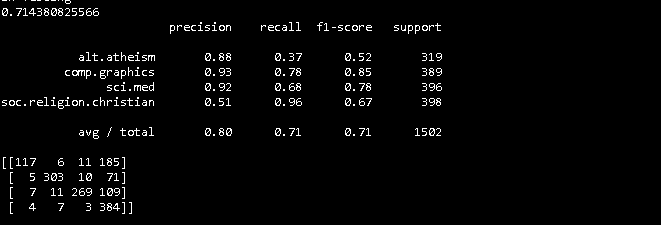
**2.1** Casual Model



**2.2 – 2.5**



**3.** Model designed only for 4 categories namely - alt.atheism, soc.religion.christian, comp.graphics, sci.med as discussed with the TA. Accuracy score, classification report and confusion matrix given below.



References

[1] 2017. [Online]. Available: http://www.ics.uci.edu/~smyth/courses/cs274/notes/EMnotes.pdf. [Accessed: 16- Oct- 2017].

[2] "sklearn.metrics.silhouette\_score — scikit-learn 0.19.0 documentation", Scikit-learn.org, 2017. [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette\_score.html. [Accessed: 16- Oct- 2017].

[3] 2017. [Online]. Available: http://web.engr.oregonstate.edu/~xfern/classes/cs534/notes/Unsupervised-model-11.pdf. [Accessed: 16- Oct- 2017].

[4] "piedenis / Lea — Bitbucket", Bitbucket.org, 2017. [Online]. Available: https://bitbucket.org/piedenis/lea. [Accessed: 16- Oct- 2017].

[5] "Bayesian network", En.wikipedia.org, 2017. [Online]. Available: https://en.wikipedia.org/wiki/Bayesian\_network. [Accessed: 16- Oct- 2017].

[6] "Working With Text Data — scikit-learn 0.19.0 documentation", Scikit-learn.org, 2017. [Online]. Available: http://scikit-learn.org/stable/tutorial/text\_analytics/working\_with\_text\_data.html. [Accessed: 16- Oct- 2017].

[7] "The Bernoulli model", Nlp.stanford.edu, 2017. [Online]. Available: https://nlp.stanford.edu/IR-book/html/htmledition/the-bernoulli-model-1.html. [Accessed: 16- Oct- 2017].

[8] "Text Classification Using Naive Bayes", YouTube, 2017. [Online]. Available: https://www.youtube.com/watch?v=EGKeC2S44Rs. [Accessed: 16- Oct- 2017].