## **CAPSTONE PROJECT PROPOSAL - Dresses Sales Recommendation**

## Q: What is the problem you want to solve?

A: Small online retailers usually set themselves apart from brick and mortar stores, traditional brand names, and giant online retailers by offering goods at exceptional value. In addition to price, they compete for shoppers’ attention via descriptive listing titles, whose effectiveness as search keywords can help drive sales. In this project, machine learning techniques will be applied to build a recommender system to build up meaningful recommendations for users about dresses that might interest them.

Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

A: The possible clients can be -

* E-commerce websites who use the recommender system developed by analysing the data set to boost the sale of the dress.
* Internet search engines for mining the dress based on the attribute tags.
* Buyers of the dresses can be benefit by the market trends to help them in selecting dress of their choice.

Q: What data are you going to use for this? How will you acquire this data?

A: The dataset for this problem is borrowed from the Center for Machine Learning and Intelligent Systems repository of University of California, Irvine. The data sets can be downloaded from this site - https://archive.ics.uci.edu/ml/datasets/Dresses\_Attribute\_Sales.

The data sets consists of –

1. Attribute DataSet.xlsx - Each row describes attributes of dresses and their recommendations according to their sales.
2. Dress Sales.xlsx – Each row represents the sale of dress on specific days.
3. URL Table.txt – Text file containing the URL’s of each dress.

These files are to be analysed to build the model. The use of the data sets 2 and 3 are not clearly understood at this point of time.

In brief, outline your approach to solving this problem (knowing that this might change later).

A: Possible models that can be developed will be based on clustering and classification algorithms.

* Clustering – Model build on K Nearest Neighbours clustering algorithm, to aid in a recommendation engine for e-commerce websites.
* Classification – Model based on logistical regression to classify the buying pattern of the dresses.

Possible challenges

* Understanding the relationship between the attributes and missing values in the datasets.
* Testing the accuracy of the model.
* Might need advanced knowledge of supervised (for classification) and unsupervised (for clustering) algorithms to build a highly efficient model.

Q: What are your deliverables?

A: The following are the intended deliverables of the project

* Code Project in R and Python scripts
* Project Document clearly outlining
* The problem statement being solved.
* Approaches used to model the problem
* Explanation of data sets and predictive models used in modelling
* Results of the predictive model
* Conclusion and further recommendations
* Business study case presentation for prospective customers
* R Shiny web application