Big Data Applications and Analytics

Home Depot Search Relevance Project

<u>Work Breakdown</u>

Team Size: 3

Team Members: Saif Syed Ahmed,   Snehil Vishwakarma,   Sruthi Mallina

Initial Task: Three of us have initially worked on the dataset of World development Indicators. However while analyzing its attributes, we understood that the dataset was more inclined towards exploration and visualization but not towards analysis so we had to scrape it off. So we requested Professor Laszewski to kindly allow us to change the dataset. Professor Laszewski approved us to use a different dataset. We explored various datasets and finalized the dataset of Home Depot Search Relevance and studied its attributes and the trends and finally decided to perform analysis on it.

Task Number	Task Name	Task Description	Team Members	Gitlab Pointer
1	Data Exploration showing Data Statistics and Frequency Distribution	Data statistics depicting mean, variance, standard deviation etc, are performed on the merged data frames. Frequency distribution of certain attributes computed and worked on clustering	Sruthi Mallina	297cfaf01a1993462047474a41792549e1b857b1
2	Data Preprocessing	The various techniques of preprocessing like stemming, common words , frequency are	Saif, Sruthi, Snehil	2937139130772346babbbd8fa505386538d4eea6
3	Data Analysis and Algorithm Implementation	Applied the feature extraction and implemented the random forest algorithm	Saif, Sruthi, Snehil	d18ce9bfd6849c5bf1cddfd4d267e0db4fc4619e
4	Data Visualization of Line chart and Histogram	Designed the linecharts and histogram on relevance and frequencies respectively	Saif, Snehil	d3c1123cebd8952b085f1fe7c4a9370eb710cae0
5	Data Visualization of Pie Chart	Piecharts depicting subranges of relevancies according to different splits	Snehil	14a0167dda187d3686e503f04e469df19bab197f
6	Data Visualization of Scatter Plot	Scatter plot analyzing the features of count of unique relevancies	Saif, Snehil	db1fc4f82cc0de0a1600d8cb456bce8cf1c0d5b8
7	Data Visualization of Word Cloud	Word cloud on search terms.	Snehil, Sruthi	5304aa351d88d526d7a006c1605e905e447f3cba