Beauty Kabore

INFO 4670

Summer 2024

**Textual Association of Netflix and Google Employee Review**

**ABOUT THE DATASET:**

The dataset contains reviews from Netflix and Google employees about their thoughts on working for these companies. The dataset was originally sourced from Kaggle.

The data contains seventeen regular attributes with data types including text, dates, location and integer.

* **Select Attributes:** ‘cons’ to determine which group of words participants in the review associate negatively with regards to working with the companies.
* **Replace Missing Values:** all integer attributes with missing values were replaced with zero. This was done by filtering the attributes using subsets.
* **Process Documents from Data**
  + **Tokenize:** used to split the text document into a sequence of tokens. The default splitting mode of ‘non letters’ was applied. This resulted in tokens consisting of one single word.
  + **Transform Case** changes all words into lower case.
  + **Filter Stop words:** used to remove stop words in the data such as ‘is’, ‘are’, etc.
  + **Filter Tokens:** used to filter tokens regarding their length. The minimal number of characters that a token must contain to be considered is 4 and the maximal number is 35.
* **FP-Growth:** This was applied to calculate frequently occurring words within the dataset: helping to determine which words are strongly associated.With regards to our selected attribute (‘con’), we will be able to see which group of words employees associate with not being happy at their various workplace. The min requirement parameter was support (to determine cutoff) and the min support was 0.05.
* **Create Association Rules:** This was created and applied to help discover relationships between attributes in the dataset. A criterion parameter confidence and a minimum confidence of 0.3 were used for the selection of the rules. A Gain of 2, and Laplace k of 1 was also applied**.**

**ANALYSIS GOAL:**

1. What attributes are employees not happy with at Netflix and Google?

**RESULTS OF ANALYSIS:**



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**CONCLUSION:**

As shown above from the graph and association rule, work, life, and balance are strongly associated with each other with high levels of confidence. We can infer that employees at Google and Netflix are not happy with their work-life balance.