Methodology

As part of the project we have to identify the Target attributes (Audio, Battery, Camera, Design, Fingerprint, Memory, OS, Performance and Others) and each target attribute has a list of vector attributes as below an example tabulated below.

|  |  |  |
| --- | --- | --- |
| Sl.no | Target Attribute | Vector Attributes |
| 1 | Audio | Calling, Volume, Breaking, Voice |
| 2 | Battery | Usage, Backup, Life, Minutes, Charge |
| 3 | Camera | Front, Rear, Resolution, Clarity |
| 4 | Design | Display, Body, Sleek |
| 5 | Fingerprint | Gesture, Scanner |
| 6 | Memory | GB, Ram, SD Card |
| 7 | OS | Update, Android, Software, Snapdragon |
| 8 | Performance | Lag, Speed, Hang |
| 9 | Others | Budget, Warranty |

Initially vector attributes were identified with the method TFIDF (Term Frequency & Inverse Document Frequency) where we calculated term frequency of a word i.e., number of occurrences of a word in a review and Document frequency is also calculated i.e. occurrence of a word in all the reviews and its inverse is taken as shown in the below formula.



Here df(w) (‘w’ indicates a word) is known as Document frequency and the result log(N/df(w)) is known as Inverse Document Frequency. In the reviews we often encounter grammatical expressions such as ‘Have’, ‘has’ or and etc. these words are encountered in all the reviews but they do not hold any significant value in terms of explaining features. Hence in order to reduce their importance we calculate their inverse resulting in reduced overall importance, the following example exemplifies the importance of TFIDF in terms of attribute identification. For demonstration purpose we are trying to prove the word ‘The’ as insignificant.

Review 1 – The quality of the camera is good.

Review 2 – The Battery backup could be better.

TFIDF value calculation for Review 1, here N is the total number of reviews. Therefore N=2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Word (w)** | | **Tf(w)** | **Df(w)** | **Log(N/df(w))** | **TFIDF=tf(w)\*log(N/df(w)** |
| Camera | 1 | | 1 | 0.6931472 | 0.6931472 |
| Good | 1 | | 1 | 0.6931472 | 0.6931472 |
| Is | 1 | | 1 | 0.6931472 | 0.6931472 |
| Of | 1 | | 1 | 0.6931472 | 0.6931472 |
| Quality | 1 | | 1 | 0.6931472 | 0.6931472 |
| The | 1 | | 2 | -0.4054651 | -0.4054651 |

If we observe the result, TFIDF value for the word ‘The’ is in negative indicating that its insignificant similarly the rest of grammatical lexicons will be identified and will be eliminated after we set the threshold TFIDF value. After identifying potential vector attributes, we manually tag them to respective targets.

**Mapping Features to a Review**

The following Cosine Similarity formula is being used to map a review to respective feature where D1 is a Document comprising single review and D2 is a document comprising attribute data of any specific feature. We have to compare the review for each feature separately to understand if the review is speaking about particular feature or not.

****

The following example where we have taken a review directly from the e-commerce site for vivo-v5 phone is taken and illustrated how similarity is being calculated.

Review - I have compared this phone with lenovo, motorola, and redmi 4GB RAM and 64 GB Storage phones and in all of them this phone is awsome..go for it without any hesitation. front camera is awsome. i am giving 5 out of 5 to this phone. battery is good. the nought OS update is already available for this phone

1. Load the review to the Document D1
2. Load any feature’s attribute vector to D2
3. Create a Document Term Matrix
4. Send the matrix to the Cosine function of LSA package in R
5. Verify if the result value lies in between 0 and 1 where 0 indicates not a single attribute related to the feature is available in the review and vice versa.

Document Matrix with respect to Camera Attributes

D1 D2

1. camera 1 1

2. clarity 1 0

3. clear 1 0

4. front 1 1

5. mega 1 0

6. pixel 1 0

7. rear 1 0

8. resolution 1 0

.. . . . . . . .

There is 0.0632772 similarity between the two documents

Document Matrix with respect to Memory Attributes

D1 D2

34. ram 1 1

35. redmi 1 0

38. them 1 0

39. this 4 0

40. to 1 0

41. update 1 0

42. with 1 0

43. without 1 0

44. card 0 1

45. inbuilt 0 1

46. internal 0 1

47. memory 0 1

48. rom 0 1

49. sd 0 1

There is 0.0949158 similarity between the two documents

Document Matrix with respect to Operating System Attributes

D1 D2

30. os 1 1

38. them 1 0

39. this 4 0

40. to 1 0

41. update 1 1

42. with 1 0

43. without 1 0

44. android 0 1

45. operating 0 1

46. software 0 1

47. system 0 1

There is 0.07749843 similarity between the two documents

Document Matrix with respect to Battery Attributes

$matrix

D1 D2

11. battery 1 1

44. backup 0 1

45. charge 0 1

46. charger 0 1

47. charging 0 1

48. heat 0 1

49. heating 0 1

50. hours 0 1

51. life 0 1

52. mah 0 1

53. minutes 0 1

54. overheating 0 1

55. usage 0 1

There is only 0.02632491 similarity between the two documents

Document Matrix with respect to Audio Attributes

$matrix

D1 D2

46. calling 0 1

47. clear 0 2

48. crystal 0 1

49. loud 0 1

50. melodious 0 1

51. music 0 2

52. network 0 1

53. output 0 1

54. sound 0 1

55. speaker 0 1

56. voice 0 1

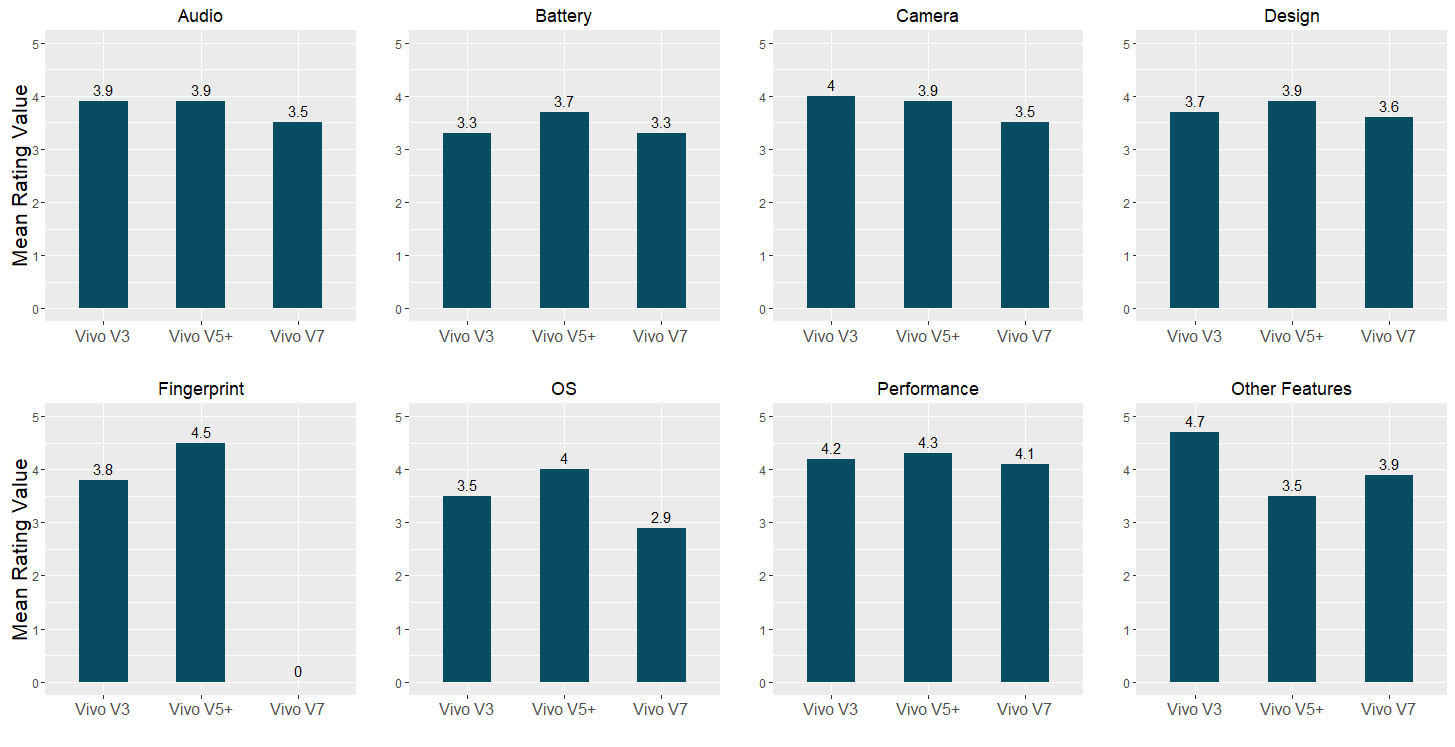
57. volume 0 1

As there are no attributes related to audio is observed in the review, we got 0 similarity.

The following highlighted attributes has been identified via this process.

Review - I have compared this phone with lenovo, motorola, and redmi 4GB RAM and 64 GB Storage phones and in all of them this phone is awsome..go for it without any hesitation. front camera is awsome. i am giving 5 out of 5 to this phone. battery is good. the nought OS update is already available for this phone

**Analysis**



After tagging the reviews with respective features, we have plotted the average of their ratings for the models Vivo V3, Vivo V5+ and Vivo V7. The following are the critical inferences made after

* Customers are experiencing quick battery drain in V7 model compared to other brands, that explains the decline in rating from V5 to V7.
* Users have experienced focus and zooming issues in V7 Camera which caused decline in the rating.
* There is sudden drop in the Operating System feature in Vivo V7 as the company failed to push regular OS updates compared to competitors.
* Fingerprint sensor is fast compared to V3 to V5.
* Overall Design and Performance have been consistent and there are few design issues like headphone jack position, grip and screen issues needs to addressed.