1

2

3

# 1

# 2

# 3

CHAPTER 4

# RESULTS AND DISCUSSION

This chapter demonstrates the results of machine learning models for aspect-based sentiment classification. We evaluated the performance of machine learning models corresponding to each aspect such as price, weight, color, service, and size. The models performed differently in different aspects of the dataset. We analyse all performances and comparison in this chapter and explain the results in detail. In the end, result and the significance of the machine learning models is discussed in detail.

As discussed in previous chapter, three reviews datasets of eBay and Amazon has been collected from different sources. On these datasets, three set of analysis has been performed to compare the different machine learning algorithms. Parameters setting have a great significance in the analysis of machine learning algorithm.

## Hyperparameters for Classifiers

The classification of the sentiment using the text is done in this study using the supervised machine learning approach. All used classifiers are discussed in the Introduction chapter, we train all these models with different hyperparameters setting and we find these hyperparameters with the hit and trials method. These classification model achieved their best results with these hyperparameters setting.

Table 4.1: Hyperparamters for classifier

|  |  |
| --- | --- |
| Classification Model | Hyperparameters |
| Random Forest | n\_estimators=268  max\_depth= 250 |
| Logistic Regression | solver=’saga’  multi\_class=’ovr’  C=3.0 |
| Support Vector Machine | Kernel=’sigmoid’ |
| K Nearest Neighbour | **n\_neighbors=100** |
| Gaussian Naïve Bayes | **Default Setting** |

## Comparative Analysis of Machine Learning Algorithms on Dataset\_1

Machine learning is performed well when significant number of data is available for training. The well-known machine algorithms are applied on dataset\_1 on same hyperparameters as mentioned above in table 4.1. The evaluation of the machine learning algorithms is given in table 4.2.

It is shown that the Support Vector Machine, Logistic Regression and Random Forest have performed better as compare to the rest of the algorithms. To achieve the optimum results, these algorithms are combined under the architecture of Voting Classifier. The results are presented in table 4.3. It is observed that the combination of Logistic Regression and Random Forest is performed well ascpmpare to the other combinations.

Table 4.2: Comparison of ML algorithms on dataset\_1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Percision | Recall | F1 Score |
| Support Vector Machine | 91.86% | 0.92 | 0.92 | 92% |
| K-Nearest Neighbor | 75.95% | 0.80 | 0.75 | 75% |
| Gaussian Naïve Bayes | 85% | 0.85 | 0.85 | 85% |
| Logistic Regression | 92.799% | 93 | 93 | 93% |
| Random Forest | 92 | 92 | 92 | 92 |
| Recurrent Neural Network | 80 | 80 | 82 | 76 |

Table 4.3: Comparison of Proposed Methodology

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Voting Classifier (Logistic Regression +Random Forest) | 93.14% | 94 | 93 | 93% |
| Voting Classifier (Random Forest + Support Vector Machine) | 92.68% | 93 | 93 | 93% |
| Voting Classifier (Logistic Regression + Support Vector Machine) | 92.33% | 92 | 92 | 92% |
| Base Paper Technique (Gini Index + SVM) | 84.66% | 85 | 85 | 85% |

## Comparative Analysis of Machine Learning Algorithms on Dataset\_2

To analyse the results, the above used machine learning algorithms are applied in dataset\_2 with the same hyperparameters mentioned in table 4.1. The evaluation of the machine learning algorithms is given in table 4.3.

Table 4.4: Comparison of ML algorithms on dataset\_2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 89% | 89 | 89 | 71% |
| K-Nearest Neighbor | 72% | 74 | 72 | 64% |
| Gaussian Naïve Bayes | 65% | 65 | 65 | 90% |
| Logistic Regression | 89.996% | 90 | 90 | 87% |
| Random Forest | 87% | 87 | 88 | 90% |
| Recurrent Neural Network | 82% | 80 | 74 | 76% |

The above table shows that the Support Vector Machine, Logistic Regression and Random Forest have performed better as compare to the rest of the algorithms. To improve the accuracy, these algorithms are combined under the architecture of Voting Classifier. The results are presented in table 4.5. It is observed that the combination of Logistic Regression and Random Forest is performed well as compare to the other combinations.

Table 4.5: Comparison of Proposed Methodology

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Voting Classifier (Logistic Regression +Random Forest) | 90.069% | 90 | 90 | 89% |
| Voting Classifier (Random Forest + Support Vector Machine) | 89.371% | 89 | 89 | 89% |
| Voting Classifier (Logistic Regression + Support Vector Machine) | 89.334% | 89 | 89 | 77% |
| Base Paper Technique (Gini Index + SVM) | 76% | 76 | 76 | 76% |

## Comparative Analysis of Machine Learning Algorithms on Aspect Level Dataset\_1

With the help of analysis performed in previous sections, it is shown that proposed methodology (combination of Logistic Regression and Random Forest) has performed well as compare to the other machine learning algorithms. On the basis of this analysis, aspect level dataset is analysed and compared with the different machine learning algorithms. The aspect level data is imbalanced. To deal with the imbalanced dataset, under-sampling and over-sampling technique is performed. In this section, machine learning algorithms and proposed methodology is applied on all three type of datasets, without any sampling dataset, under-sampling datasets and over-sampling datasets.

### Without Imbalance Data

This section contains the comparison of different machine learning algorithms on aspect level dataset. Aspect related data set was extracted from the large dataset. Extracted data set is imbalanced because number of positive reviews are greater than the negative and neutral reviews. In the following subsections, machine learning algorithms have been implemented on each of the aspect’s dataset (Size, Color, Price, Service, and Weight).

#### Size

Comparison among the machine learning algorithms have been performed on size related aspect. The data set contains 6969 reviews in which 5262 reviews are positive, 1358 reviews are negative and 345 are neutral reviews. Table 4.1 shows that SVM obtained highest accuracy 85% and F1 Score as compare to other algorithms on high imbalanced dataset. Moreover, Logistic Regression also obtained the 83% accuracy while proposed technology did not provide the considerable results on imbalanced dataset.

Table 4.6: Comparison of ML Algorithms on Size Aspect on Imbalanced Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.85 | 0.84 | 0.59 | 0.64 |
| K-Nearest Neighbor | 0.75 | 0.58 | 0.34 | 0.29 |
| Random Forest | 0.78 | 0.82 | 0.45 | 0.49 |
| Logistic Regression | 0.83 | 0.89 | 0.53 | 0.59 |
| Gaussian Naïve Bayes | 0.61 | 0.64 | 0.45 | 0.49 |
| Recurrent Neural Network | 0.82 | 0.60 | 0.54 | 0.57 |
| Voting Classifier | 0.80 | 0.92 | 0.48 | 0.53 |

#### Color

In this section, comparison among the machine learning algorithms have been performed on color related aspect. The data set contains 3074 reviews in which 2239 reviews are positive, 680 reviews are negative and 153 are neutral reviews. Table 4.2 shows that SVM achieved highest accuracy 80% and F1 as compare to other algorithms on imbalanced dataset. Logistic Regression obtained the 79% accuracy and RNN achieved 78%. The proposed technology did not provide the considerable results on imbalanced dataset with respect to accuracy but achieved the highest precision value.

Table 4.7: Comparison of ML Algorithms on Color Aspect on Imbalanced Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.80 | 0.73 | 0.50 | 0.53 |
| K-Nearest Neighbor | 0.72 | 0.57 | 0.34 | 0.28 |
| Random Forest | 0.75 | 0.80 | 0.43 | 0.46 |
| Logistic Regression | 0.79 | 0.88 | 0.47 | 0.51 |
| Gaussian Naïve Bayes | 0.65 | 0.64 | 0.43 | 0.46 |
| Recurrent Neural Network | 0.78 | 0.51 | 0.44 | 0.44 |
| Voting Classifier | 0.76 | 0.90 | 0.43 | 0.45 |

#### Price

In this section, comparison among the machine learning algorithms have been performed on Price related aspect. The data set contains 32139 reviews in which 24105 reviews are positive, 6592 reviews are negative and 1417 are neutral reviews. Table 4.3 shows that SVM achieved highest accuracy 91% as compare to other algorithms on large and imbalanced dataset. Moreover, Logistic Regression also obtained the 90% accuracy and RNN achieved 90% while proposed methodology also provide the considerable results on large and imbalanced dataset in terms of accuracy and precision.

Table 4.8: Comparison of ML Algorithms on Price Aspect on Imbalanced Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.91 | 0.88 | 0.76 | 0.80 |
| K-Nearest Neighbor | 0.65 | 0.52 | 0.46 | 0.35 |
| Random Forest | 0.80 | 0.92 | 0.44 | 0.48 |
| Logistic Regression | 0.90 | 0.90 | 0.69 | 0.75 |
| Gaussian Naïve Bayes | 0.50 | 0.44 | 0.49 | 0.42 |
| Recurrent Neural Network | 0.90 | 0.81 | 0.76 | 0.78 |
| Voting Classifier | 0.87 | 0.92 | 0.58 | 0.65 |

#### Service

In this section, comparison among the machine learning algorithms have been performed on Service related aspect. The data set contains 6901 reviews in which 5023 reviews are positive, 1515 reviews are negative and 360 are neutral reviews. Table 4.4 shows that SVM achieved highest accuracy 84% as compare to other algorithms on imbalanced dataset. Moreover, Logistic Regression also obtained the 83% accuracy. RNN and proposed methodology achieved the same accuracy of 81% but precision value of voting classifier is greater than the RNN.

Table 4.9: Comparison of ML Algorithms on Service Aspect on Imbalanced Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.84 | 0.79 | 0.59 | 0.64 |
| K-Nearest Neighbor | 0.74 | 0.58 | 0.48 | 0.46 |
| Random Forest | 0.77 | 0.83 | 0.43 | 0.46 |
| Logistic Regression | 0.83 | 0.88 | 0.53 | 0.58 |
| Gaussian Naïve Bayes | 0.64 | 0.54 | 0.45 | 0.47 |
| Recurrent Neural Network | 0.81 | 0.64 | 0.53 | 0.58 |
| Voting Classifier | 0.81 | 0.88 | 0.51 | 0.56 |

#### Weight

In this section, comparison among the machine learning algorithms have been performed on Weight related aspect. The data set contains 2068 reviews in which 1524 reviews are positive, 467 reviews are negative and 75 are neutral reviews. Table 4.5 shows that Support Vector Machine and Logistic Regression achieved highest accuracy 79% as compare to other algorithms on imbalanced dataset. Moreover, Random Forest and Voting Classifier achieved the same accuracy, recall and precision but F1 score of voting score algorithms are greater than the Random Forest.

Table 4.10: Comparison of ML Algorithms on Weight Aspect on Imbalanced Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.79 | 0.49 | 0.41 | 0.42 |
| K-Nearest Neighbor | 0.72 | 0.54 | 0.39 | 0.41 |
| Random Forest | 0.77 | 0.52 | 0.36 | 0.34 |
| Logistic Regression | 0.79 | 0.52 | 0.39 | 0.39 |
| Gaussian Naïve Bayes | 0.68 | 0.33 | 0.34 | 0.33 |
| Recurrent Neural Network | 0.74 | 0.71 | 0.27 | 0.40 |
| Voting Classifier | | 0.77 | 0.52 | 0.36 | 0.35 |

### With Under Sampling

This section also contains the comparison of different machine learning algorithms on aspect level dataset. Extracted data set was imbalanced because number of positive reviews are greater than the negative and neutral reviews. To resolve data unbalancing problem, under sampling technique is applied on all aspects related datasets. In the following subsections, machine learning algorithms have been implemented on each of the aspect’s dataset (Size, Color, Price, Service, and Weight).

#### Size

Comparison among the machine learning algorithms have been performed on size related aspect. The data set contains 1035 reviews in which 345 reviews are positive, 345 reviews are negative and 345 are neutral reviews. Table 4.6 shows that Logistic Regression obtained highest accuracy 83% and F1 Score as compare to other algorithms on high imbalanced dataset. Moreover, SVM and Voting Classifier achieved 81%. This shows that proposed methodology improves their results on balanced data.

Table 4.11: Comparison of Machine Learning Algorithms on Size Aspect

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.81 | 0.81 | 0.80 | 0.80 |
| K-Nearest Neighbor | 0.67 | 0.65 | 0.64 | 0.64 |
| Random Forest | 0.77 | 0.79 | 0.76 | 0.75 |
| Logistic Regression | 0.83 | 0.82 | 0.82 | 0.82 |
| Gaussian Naïve Bayes | 0.57 | 0.66 | 0.57 | 0.54 |
| Recurrent Neural Network | 0.44 | 1.00 | 0.05 | 0.10 |
| Voting Classifier | 0.81 | 0.81 | 0.80 | 0.80 |

#### Color

In this section, comparison among the machine learning algorithms have been performed on color related aspect. The data set contains 459 reviews in which 153 reviews are positive, 153 reviews are negative and 153 are neutral reviews. Table 4.7 shows that Logistic Regression achieved highest accuracy 73% and F1 as compare to other algorithms on small and balanced dataset. Moreover, SVM and KNN also achieved better accuracy as compare to rest of the mentioned algorithms. The proposed technology did not provide the considerable results on small dataset.

Table 4.12: Comparison of Machine Learning Algorithms on Color Aspect

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.72 | 0.73 | 0.73 | 0.72 |
| K-Nearest Neighbor | 0.72 | 0.68 | 0.67 | 0.67 |
| Random Forest | 0.65 | 0.74 | 0.70 | 0.65 |
| Logistic Regression | 0.73 | 0.74 | 0.74 | 0.73 |
| Gaussian Naïve Bayes | 0.62 | 0.64 | 0.60 | 0.61 |
| Recurrent Neural Network | 0.63 | 0.59 | 0.57 | 0.58 |
| Voting Classifier | 0.67 | 0.71 | 0.71 | 0.68 |

#### Price

In this section, comparison among the machine learning algorithms have been performed on Price related aspect. The data set contains 4251 reviews in which 1417 reviews are positive, 1417 reviews are negative and 1417 are neutral reviews. Table 4.8 shows that SVM and Voting Classifier achieved the highest accuracy 90% as compare to other algorithms on large and balanced dataset. Moreover, KNN did performed well because of large dataset.

Table 4.13: Comparison of Machine Learning Algorithms on Price Aspect

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.90 | 0.90 | 0.90 | 0.90 |
| K-Nearest Neighbor | 0.54 | 0.71 | 0.53 | 0.47 |
| Random Forest | 0.89 | 0.89 | 0.89 | 0.89 |
| Logistic Regression | 0.88 | 0.89 | 0.88 | 0.88 |
| Gaussian Naïve Bayes | 0.64 | 0.67 | 0.64 | 0.62 |
| Recurrent Neural Network | 0.58 | 0.71 | 0.42 | 0.53 |
| Voting Classifier | 0.90 | 0.90 | 0.90 | 0.90 |

#### Service

In this section, comparison among the machine learning algorithms have been performed on Service-related aspect. The data set contains 1080 reviews in which 360 reviews are positive, 360 reviews are negative and 360 are neutral reviews. Table 4.9 shows that SVM and Voting Classifier achieved highest accuracy and other evaluation measures as compare to other algorithms on balanced dataset. Moreover, Random Forest also obtained the 81% accuracy and KNN performance is also increased as compare to table 4.8 because data size is decreased.

Table 4.14: Comparison of Machine Learning Algorithms on Service Aspect

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.82 | 0.82 | 0.82 | 0.82 |
| K-Nearest Neighbor | 0.68 | 0.66 | 0.68 | 0.66 |
| Random Forest | 0.81 | 0.81 | 0.81 | 0.80 |
| Logistic Regression | 0.74 | 0.75 | 0.75 | 0.74 |
| Gaussian Naïve Bayes | 0.65 | 0.70 | 0.65 | 0.64 |
| Recurrent Neural Network | 0.72 | 0.72 | 0.60 | 0.65 |
| Voting Classifier | 0.82 | 0.82 | 0.82 | 0.81 |

#### Weight

In this section, comparison among the machine learning algorithms have been performed on Weight related aspect. The data set contains 225 reviews in which 75 reviews are positive, 75 reviews are negative and 75 are neutral reviews. Table 4.10 shows that Random Forest achieved highest accuracy 69% as compare to other algorithms on small and balanced dataset. Moreover, Voting Classifier also achieved better results of accuracy.

Table 4.15: Comparison of Machine Learning Algorithms on Weight Aspect

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.60 | 0.61 | 0.60 | 0.61 |
| K-Nearest Neighbor | 0.42 | 0.44 | 0.45 | 0.43 |
| Random Forest | 0.69 | 0.77 | 0.68 | 0.64 |
| Logistic Regression | 0.60 | 0.62 | 0.61 | 0.61 |
| Gaussian Naïve Bayes | 0.44 | 0.53 | 0.48 | 0.43 |
| Recurrent Neural Network | 0.44 | 0.42 | 0.42 | 0.39 |
| Voting Classifier | 0.64 | 0.64 | 0.64 | 0.64 |

### With Over Sampling

This section also contains the comparison of different machine learning algorithms on aspect level dataset. Extracted data set was imbalanced because number of positive reviews are greater than the negative and neutral reviews. To resolve data unbalancing problem, over sampling technique SMOTE (**Synthetic Minority Oversampling Technique**) is applied on all aspects related datasets. The SMOTE technique generates the negative and neutral reviews to balance the count of reviews. In the following subsections, machine learning algorithms have been implemented on each of the aspect’s dataset (Size, Color, Price, Service, and Weight).

#### Size

Comparison among the machine learning algorithms have been performed on size related aspect. The data set contains 15786 reviews in which 5262 reviews are positive, 5262 reviews are negative and 5262 are neutral reviews. Table 4.11 shows that SVM, Random Forest and Voting Classifier obtained highest accuracy 96% and F1 Score as compare to other algorithms on large and balanced dataset. Moreover, Logistic Regression also obtained the 95% accuracy while KNN did not provide the considerable results on large dataset.

Table 4.16: Comparison of ML Algorithms on Size Aspect on Balanced Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.96 | 0.96 | 0.96 | 0.96 |
| K-Nearest Neighbor | 0.35 | 0.45 | 0.34 | 0.18 |
| Random Forest | 0.96 | 0.96 | 0.96 | 0.96 |
| Logistic Regression | 0.95 | 0.95 | 0.95 | 0.95 |
| Gaussian Naïve Bayes | 0.88 | 0.90 | 0.88 | 0.87 |
| Recurrent Neural Network | 0.72 | 0.65 | 0.60 | 0.62 |
| Voting Classifier | 0.96 | 0.96 | 0.96 | 0.96 |

#### Color

In this section, comparison among the machine learning algorithms have been performed on color related aspect. The data set contains 6717 reviews in which 2239 reviews are positive, 2239 reviews are negative and 2239 are neutral reviews. Table 4.12 shows that SVM, Random Forest and Voting Classifier obtained highest accuracy 96% and F1 Score as compare to other algorithms on large and balanced dataset. Moreover, Logistic Regression also obtained the 94% accuracy while KNN did not provide the considerable results on large dataset.

Table 4.17: Comparison of ML Algorithms on Color Aspect on Balanced Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.96 | 0.96 | 0.96 | 0.96 |
| K-Nearest Neighbor | 0.61 | 0.82 | 0.63 | 0.52 |
| Random Forest | 0.96 | 0.96 | 0.96 | 0.96 |
| Logistic Regression | 0.94 | 0.94 | 0.94 | 0.94 |
| Gaussian Naïve Bayes | 0.91 | 0.93 | 0.91 | 0.91 |
| Recurrent Neural Network | 0.67 | 0.57 | 0.59 | 0.58 |
| Voting Classifier | 0.96 | 0.96 | 0.96 | 0.96 |

#### Price

In this section, comparison among the machine learning algorithms have been performed on Price related aspect. The data set contains 72315 reviews in which 24105 reviews are positive, 24105 reviews are negative and 24105 are neutral reviews. Table 4.13 shows that Voting Classifier and Random Forest achieved highest accuracy 97% as compare to other algorithms on large and balanced dataset. Moreover, Logistic Regression also obtained the 96% accuracy while KNN did not provide the considerable results on large dataset.

Table 4.18: Comparison of ML Algorithms on Price Aspect on Balanced Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.96 | 0.96 | 0.96 | 0.96 |
| K-Nearest Neighbor | 0.41 | 0.77 | 0.40 | 0.30 |
| Random Forest | 0.97 | 0.97 | 0.97 | 0.97 |
| Logistic Regression | 0.96 | 0.96 | 0.96 | 0.96 |
| Gaussian Naïve Bayes | 0.89 | 0.88 | 0.89 | 0.88 |
| Recurrent Neural Network | 0.78 | 0.75 | 0.66 | 0.70 |
| Voting Classifier | 0.97 | 0.97 | 0.97 | 0.97 |

#### Service

In this section, comparison among the machine learning algorithms have been performed on Service-related aspect. The data set contains 15069 reviews in which 5023 reviews are positive, 5023 reviews are negative and 5023 are neutral reviews. Table 4.14 shows that Voting Classifier, Random Forest and Support Vector Machine achieved highest accuracy 96% as compare to other algorithms on large and balanced dataset. Moreover, Logistic Regression also obtained the 95% accuracy while KNN did not provide the considerable results on large dataset.

Table 4.19: Comparison of ML Algorithms on Service Aspect on Balanced Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.96 | 0.96 | 0.96 | 0.96 |
| K-Nearest Neighbor | 0.62 | 0.82 | 0.62 | 0.52 |
| Random Forest | 0.96 | 0.96 | 0.96 | 0.96 |
| Logistic Regression | 0.95 | 0.95 | 0.95 | 0.95 |
| Gaussian Naïve Bayes | 0.88 | 0.90 | 0.89 | 0.88 |
| Recurrent Neural Network | 0.72 | 0.72 | 0.60 | 0.65 |
| Voting Classifier | 0.96 | 0.96 | 0.96 | 0.96 |

#### Weight

In this section, comparison among the machine learning algorithms have been performed on Weight related aspect. The data set contains 2068 reviews in which 1524 reviews are positive, 467 reviews are negative and 75 are neutral reviews. Table 4.15 shows that Voting Classifier, Random Forest and SVM achieved highest accuracy 96% as compare to other algorithms on large and balanced dataset. Moreover, Logistic Regression also obtained the 95% accuracy while KNN did not provide the considerable results on large dataset.

Table 4.20: Comparison of ML Algorithms on Weight Aspect on Balanced Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1 Score |
| Support Vector Machine | 0.96 | 0.96 | 0.96 | 0.95 |
| K-Nearest Neighbor | 0.65 | 0.77 | 0.67 | 0.53 |
| Random Forest | 0.96 | 0.96 | 0.96 | 0.96 |
| Logistic Regression | 0.95 | 0.95 | 0.95 | 0.95 |
| Gaussian Naïve Bayes | 0.92 | 0.93 | 0.93 | 0.92 |
| Recurrent Neural Network | 0.69 | 0.62 | 0.49 | 0.55 |
| Voting Classifier | 0.96 | 0.96 | 0.96 | 0.96 |

## Analysis of Machine Learning Models

Figure 4.1 shows the results obtained with machine learning models on variety of aspects of review dataset without integration of data sampling technique. The figure reveals that Support Vector Machine has outperformed other models in terms of all aspects including size aspect, price aspect, color aspect, and service aspect. Whereas in weight aspect LR and Support Vector Machine are on the same panel. Support Vector Machine being more effective in high dimensional data produced better results. Whereas in terms of size aspect, price aspect, color aspect, and service aspect Gaussian Naive Bayes performed poorly and in weight aspect GNB along with KNN poorly performed in aspect-based sentiment analysis.

Figure 4.2 reveals the results obtained with machine learning models on a number of aspects of the review dataset with under sampling. In under sampling, there is a variation among the performance of the models in various aspects of the dataset. For instance, in the size aspect, RF outperformed other models whereas, in the color aspect

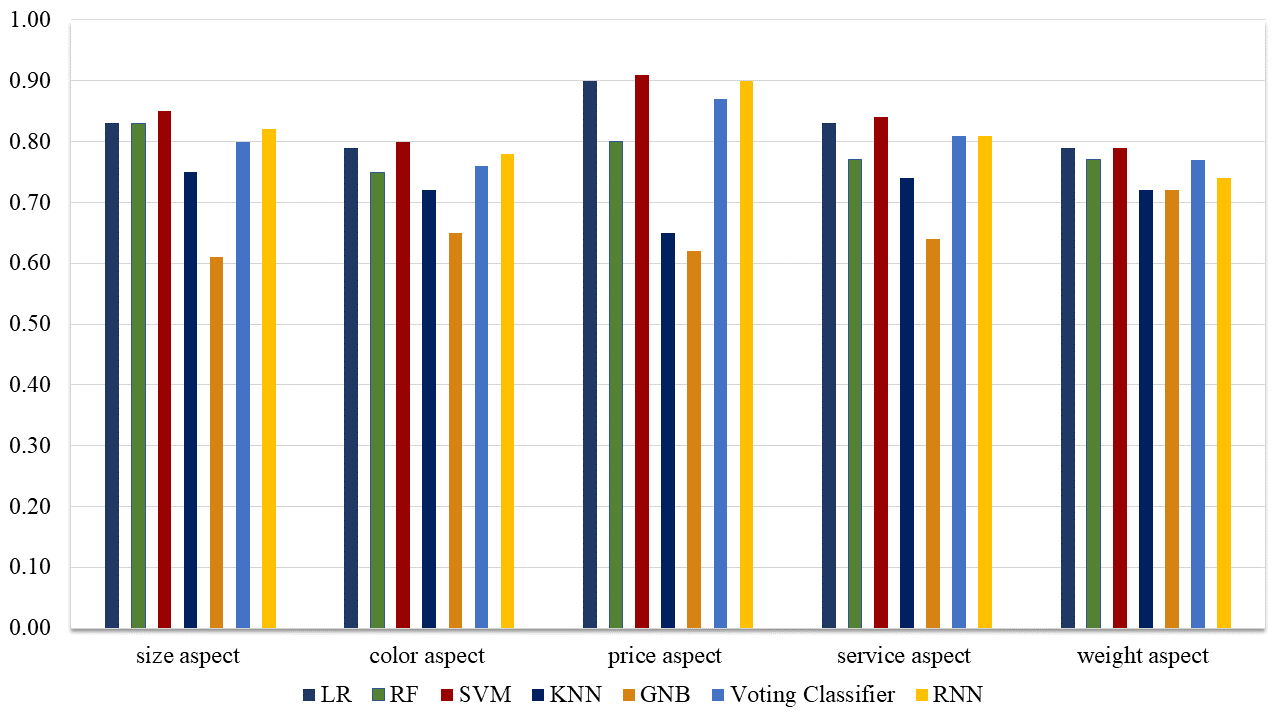


Figure 4.1 Performance evaluation results of machine learning models without any sampling technique.

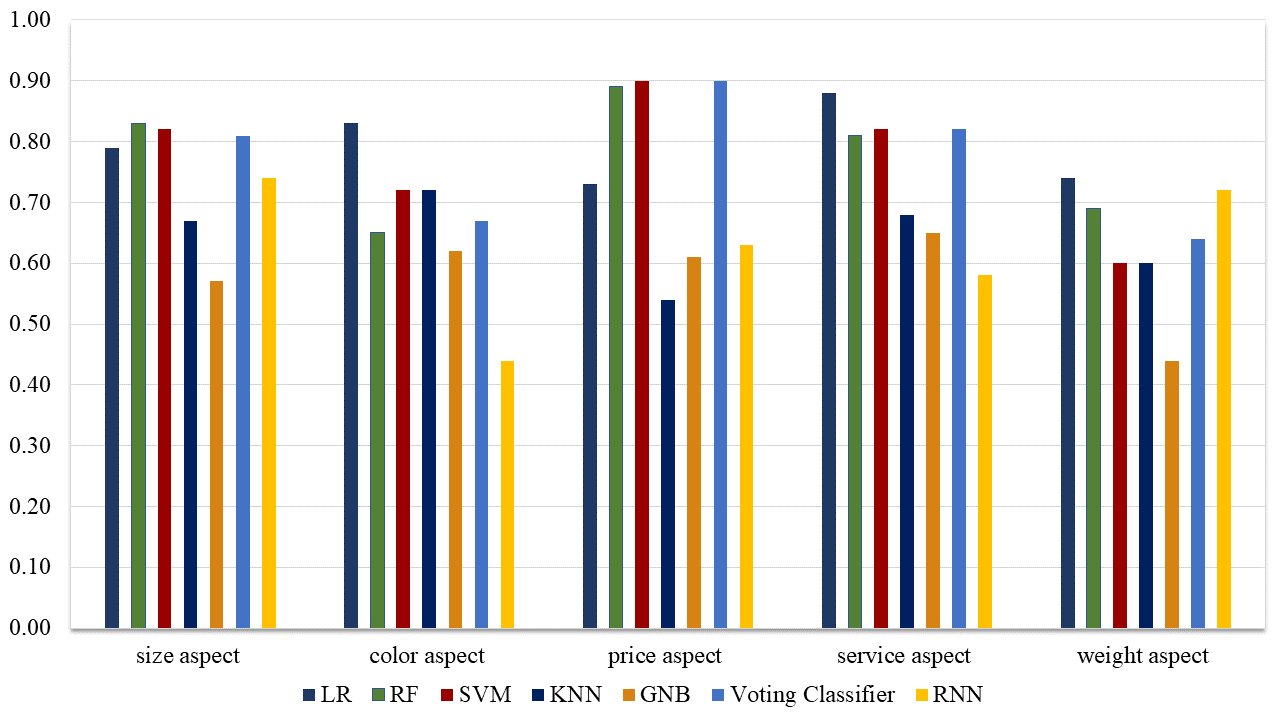


Figure 4.2 Performance evaluation results of machine learning models in aspect-based sentiment analysis with undersampling

LR outperformed other models since the color aspect has a greater number of data instances which is beneficial for LR. In case of the price aspect, SVM has shown to be the best machine learning model along with the service aspect. Whereas, in the weight aspect, both Support Vector Machine and LR have shown better performance in comparison to other models and GNB has shown poor performance.

Figure 4.3 demonstrates the results obtained with machine learning models on several aspects of the review dataset with oversampling. In the case of oversampling, RF and Support Vector Machine along with voting classifiers have been shown to outperform other machine learning models. The voting classifier gives its best performance as compared to all other models even in an imbalanced dataset case also shows the significance of the proposed ensemble models and shows that the ensemble multiple models can generate more strong results as compare to all other individual models. Whereas, KNN has poorly performed in terms of oversampling.

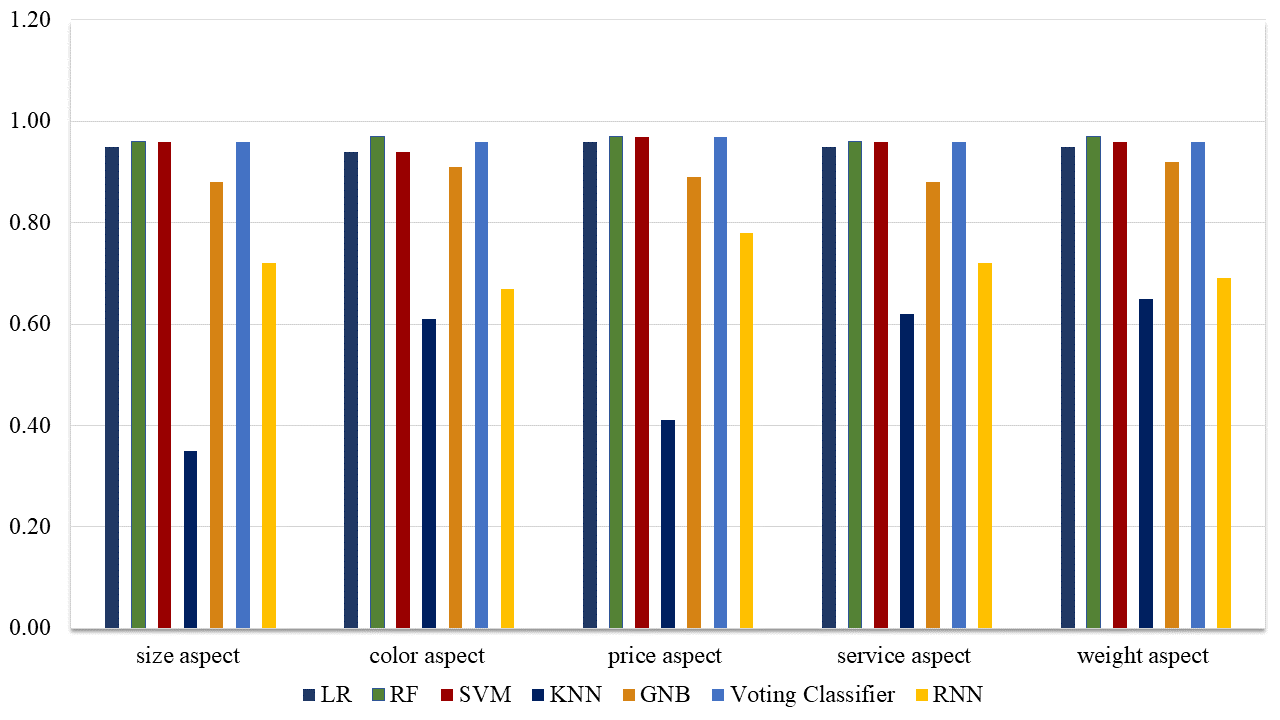


Figure 4.3 Performance evaluation results of machine learning models in aspect-based sentiment analysis with oversampling

CHAPTER 5

# CONCLUSION

This study is about the aspect level sentiment analysis using a machine learning approach. For this study, we have extracted the dataset from eBay and Amazon, and extracted dataset was imbalanced which causes models overfitting. To reduce this model's overfitting problem we have used data resampling techniques such as **Synthetic Minority Oversampling Technique (SMOTE)** and random under-sampling. The performance of all machine learning models was analyzed with a different scenario and concluded that the machine learning model's performance depends on the nature of the dataset. First, we concluded that the, if data is imbalanced models, can be overfitted on majority class data, and data resampling is the best way to handle this problem. In comparison to under-sampling, the results of over-sampling are more significant which shows the impact of SMOTE efficiency for results improvement on the imbalanced dataset. Second this we concluded that the size of data highly impact the models' performance such as the aspect have fewer data decrease the model's performance and the reason being that the fewer data generate fewer features which may not enough for the learning of models that the reason models show best results (0.97) on prices aspect because it has a large feature set. The third thing which we concluded that the tree-based models are better as compared to the linear modes when data is not linearly separable as in our case and especially LR and SVM are best performers when data is linearly separable and features set is large. We also used deep learning models with RNN architecture to perform this sentiment classification task and deep learning model results are not strong as machine learning because the used dataset is not enough for the training of deep learning models. In future work, we will generate more data to improve the performance of learning models will apply some complex deep learning approaches to improve the results.