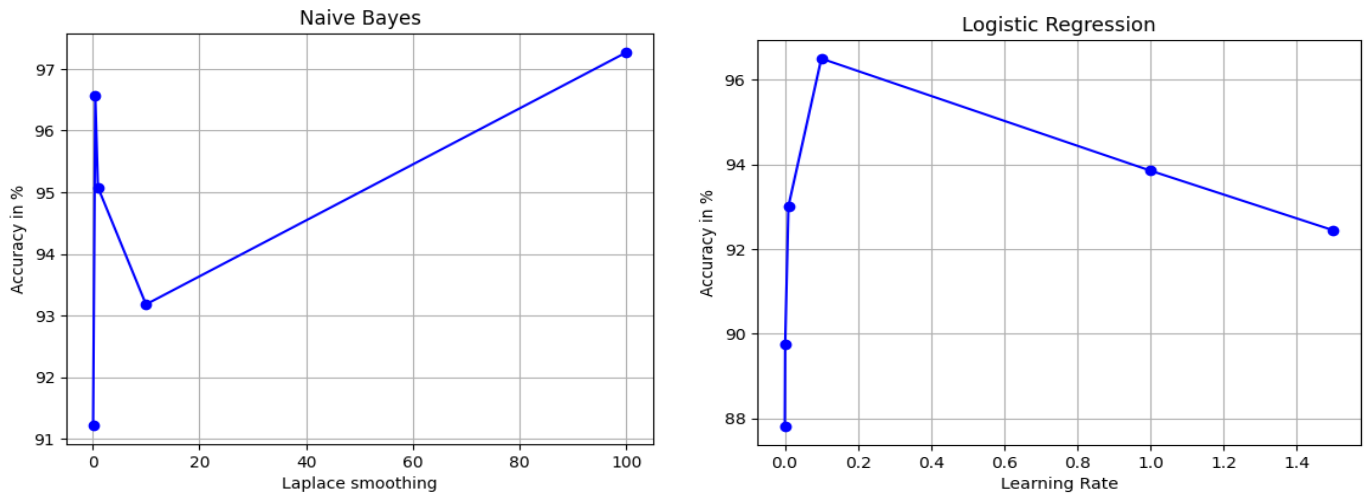
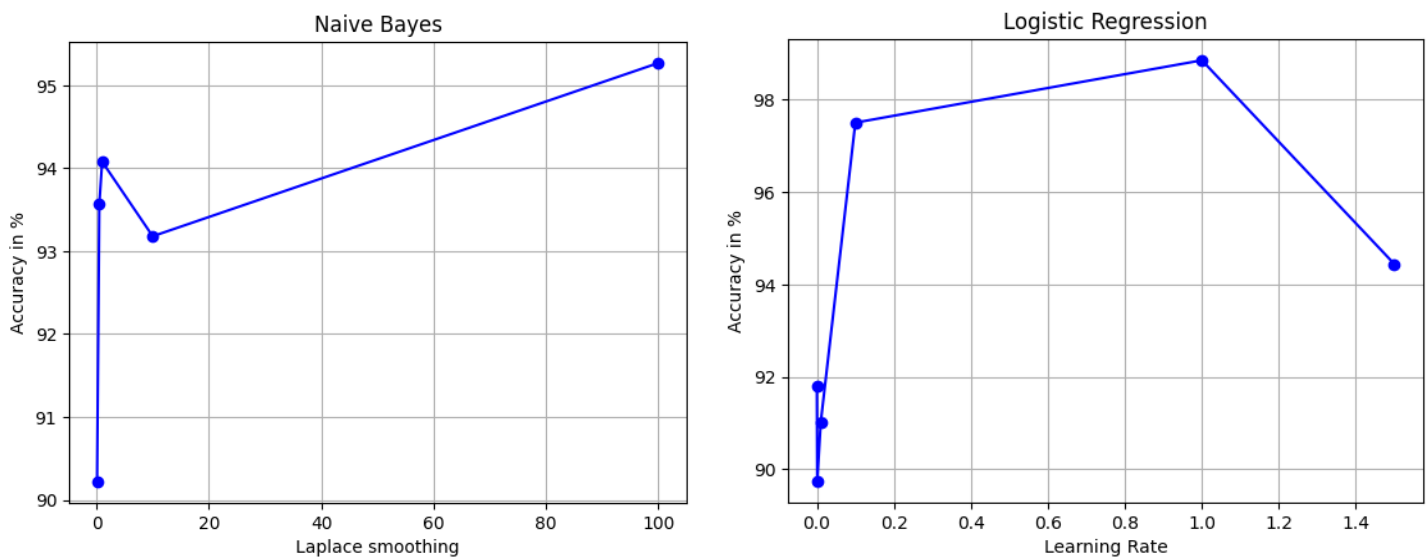


## Observation:

Results obtained using the bag of words approach for feature extraction:



Results obtained using TF-IDF approach for feature extraction:



## Conclusion:

The performance of two popular machine learning algorithms, Naive Bayes and Logistic Regression, was evaluated on the BBC dataset using two different feature extraction methods: TF-IDF and Bag of Words. The objective was to assess the accuracy of each algorithm under different parameter settings for each feature extraction method.

Firstly, when using the TF-IDF feature extraction method, Naive Bayes achieved high accuracy ranging from 91.22% to 97.27% with different Laplace smoothing parameters. On the other hand, Logistic Regression achieved accuracy ranging from 87.81% to 96.5% with different learning rates. Overall, both algorithms demonstrated commendable performance, with Naive Bayes achieving slightly higher accuracy in most cases.

Secondly, when utilizing the Bag of Words feature extraction method, Naive Bayes attained accuracy ranging from 90.22% to 95.27% with different Laplace smoothing parameters. In contrast, Logistic Regression achieved accuracy ranging from 89.74% to 98.85% with different learning rates. Notably, Logistic Regression outperformed Naive Bayes in terms of accuracy for the majority of the tested parameter values.

Comparing the two feature extraction methods, it can be concluded that both TF-IDF and Bag of Words yielded effective representations of the BBC dataset. The TF-IDF approach considered the importance of each word in the documents, while the Bag of Words method focused on the occurrence of individual words, disregarding word order and context. Despite their differences, both methods enabled Naive Bayes and Logistic Regression to capture meaningful information and achieve high accuracy.

In terms of algorithm performance, Logistic Regression consistently outperformed Naive Bayes across both feature extraction methods. Logistic Regression's ability to model complex relationships between features, leveraging the coefficients assigned to each word, contributed to its superior performance. Although Naive Bayes demonstrated competitive accuracy, Logistic Regression exhibited better adaptability and captured more nuanced patterns within the dataset.

In conclusion, both TF-IDF and Bag of Words feature extraction methods proved effective for the BBC dataset. Logistic Regression showcased stronger performance than Naive Bayes in terms of accuracy, regardless of the feature extraction approach employed. The choice of feature extraction method and algorithm should be based on the specific characteristics of the dataset and the desired trade-offs between accuracy and interpretability. Further analysis and experimentation can provide deeper insights into the performance of these algorithms on various datasets and shed light on their suitability for different natural language processing tasks.