Naive Bayes Documentation

1. Introduction

Naive Bayes is a supervised machine learning algorithm based on Bayes' Theorem. It is commonly used for classification tasks, particularly for text classification like spam detection and sentiment analysis.

2. Objective

- Develop an interactive web application using Streamlit for Naive Bayes classification.
- Provide users with the ability to upload a dataset, select features, train the model, and evaluate its performance.

3. Dataset Description

- Dataset: **A suitable dataset**
- Features: Describe the input variables
- Target: Describe the variable to predict
- Number of records and attributes

4. Implementation Details

- **Frontend**: Developed using Streamlit for user interaction.
- **Backend**: Naive Bayes implemented using `scikit-learn`.
- **Steps**:
  1. Upload the dataset in CSV format.
  2. Select features and target variables.
  3. Choose the type of Naive Bayes (Gaussian, Multinomial, or Bernoulli).
  4. Train the model using the chosen algorithm.

5. Evaluate the model using accuracy, confusion matrix, and classification report.

5. Results and Analysis

- Provide model accuracy and other evaluation metrics.
- Visualize performance using confusion matrices and other plots.
- Discuss factors influencing model accuracy.

6. Challenges and Solutions

- Managed class imbalance using appropriate preprocessing.
- Improved accuracy using feature scaling and selection.
- Addressed overfitting with proper model tuning.

7. Conclusion

Naive Bayes is a simple yet effective classification algorithm, particularly suitable for large text datasets. The interactive application allows users to explore the model's behavior and analyze its performance.

8. References

- Scikit-learn Documentation
- Streamlit Documentation
- Dataset Source