1.MNIST picture classification

We have used the following three ways of feature extraction methods

A. The general approach

This approach uses every pixel of a given picture (number picture in the case of the MNIST data set) as a feature. It then calculate the mean and the standard deviation of each pixel value and uses this data to predict the classes of pictures it's given during testing.

B. Principal Component Analysis (PCA) approach

It is a dimensionality reduction technique commonly used for feature extraction. It aims to transform a high-dimensional dataset into a lower-dimensional space while preserving the most important information and minimizing the loss of variance.

C. LPB

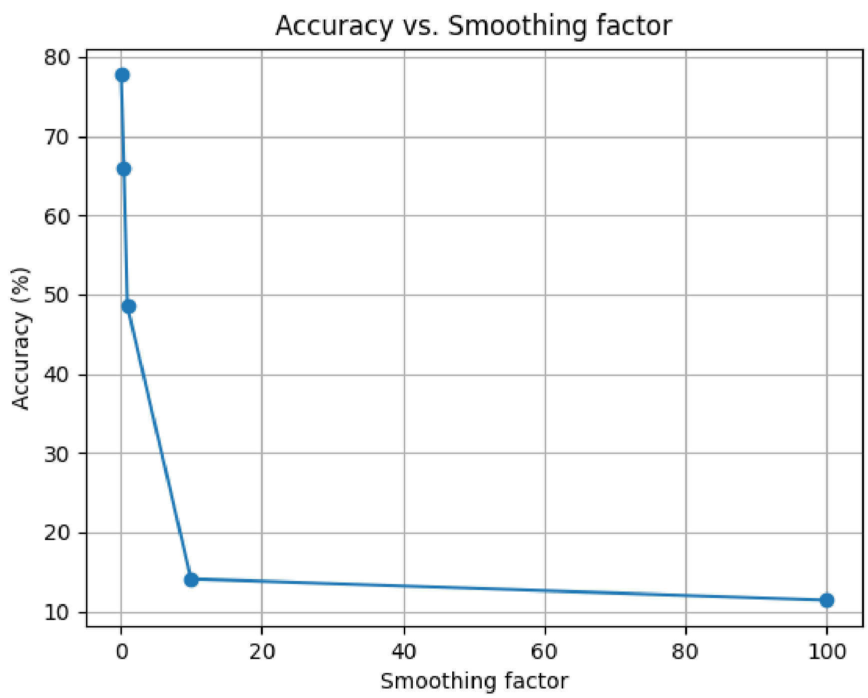
The Local Binary Pattern (LBP) feature extraction method is a technique used to describe texture patterns within an image. It works by comparing the intensity values of pixels in a circular neighborhood around each pixel and generating a binary code that represents the local texture pattern.
Sadly we weren't able to effectively implement this approach as the algorthim was taking tremendous amount of time and it was only giving as result when the dataset was very small.

Analysis of the classifiers

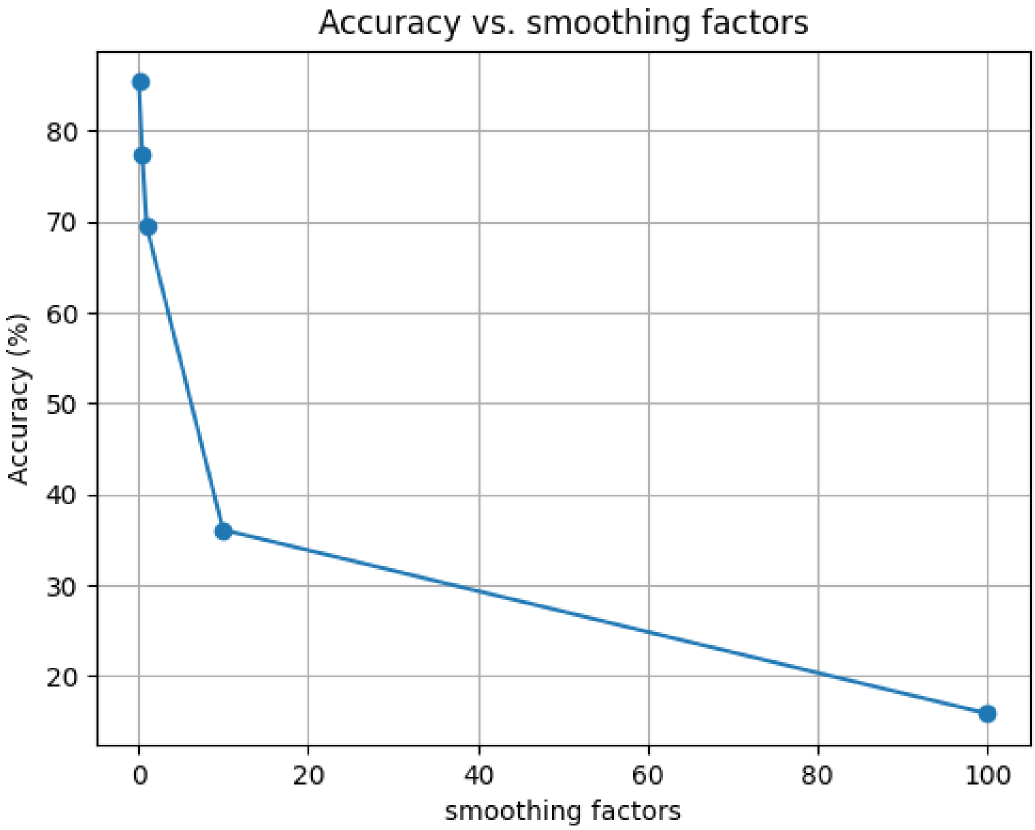
Graph plots

A. Naive Bayes Classifier

1.General approach

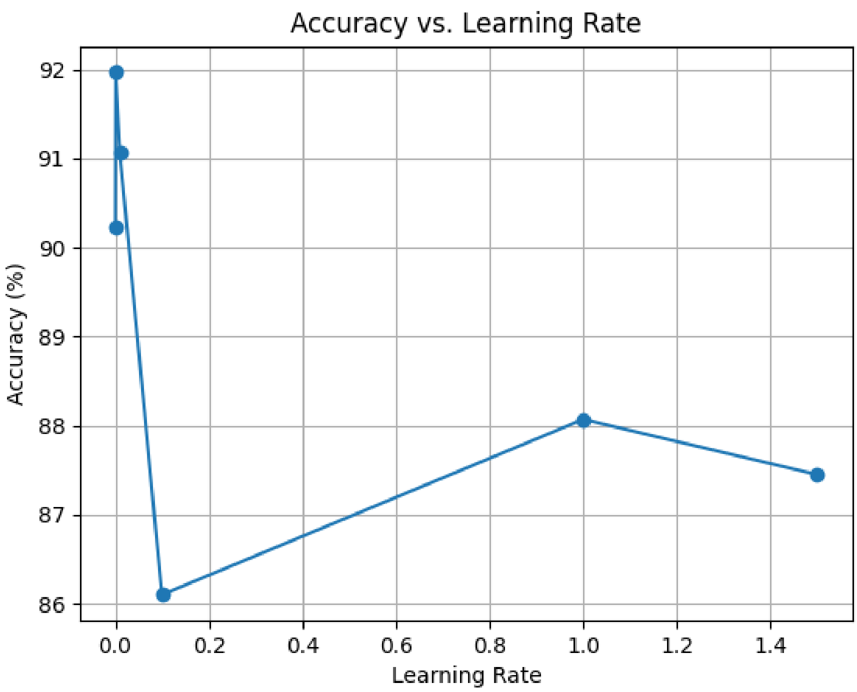


2.Pca approach

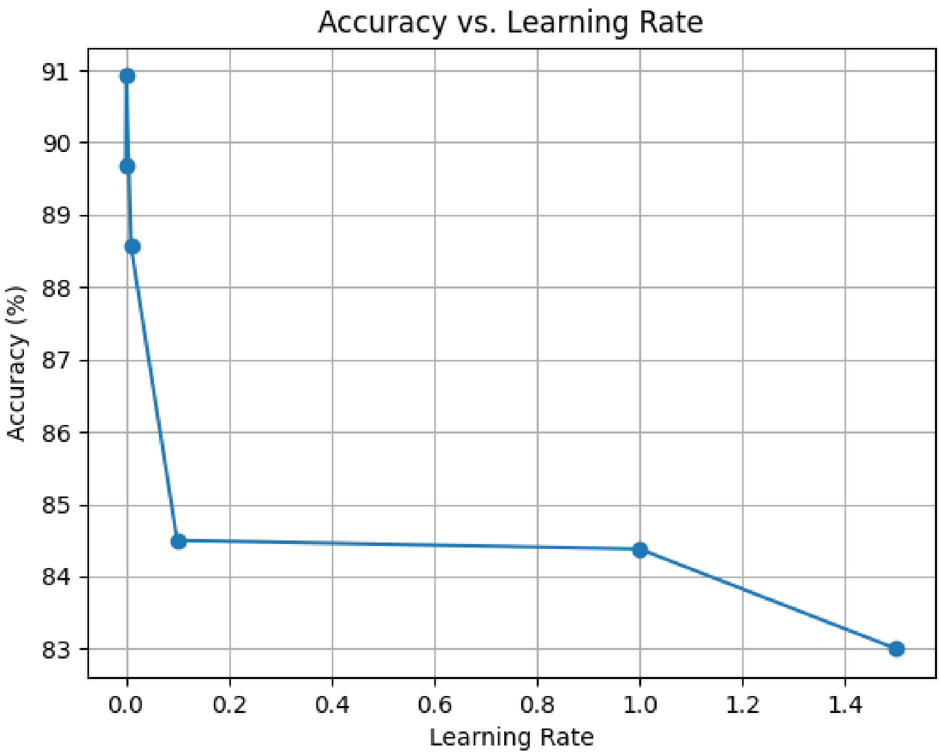


B. Logistic Regression

1.General approach



2.pca appraoch



2. BBC Text Classification

We have used the following three ways of feature extraction methods

A. Bag of Words (BoW) Feature Extraction:

The Bag of Words approach represents text data by creating a vocabulary of unique words present in the dataset. Each document or text is then represented as a vector, where each element of the vector corresponds to the frequency or presence of a specific word in the document. This method is commonly used in text classification tasks because it captures the distribution of words and their frequencies, providing a simple and efficient representation of text data.

B. TD-IDF (Term Frequency-Inverse Document Frequency) Feature Extraction:

TD-IDF is a widely used technique in text mining and information retrieval. It takes into account both the frequency of a term in a document (term frequency) and the rarity of the term across all documents (inverse document frequency). This feature extraction method assigns higher weights to terms that are frequent in a document but rare in the entire corpus, thus capturing the importance of terms in a specific document while down-weighting commonly occurring terms. TD-IDF is effective in capturing the discriminative power of terms in classification tasks.

C. **Term Presence Feature Extraction:**

Term Presence represents whether a particular term is present or absent in a document, rather than considering the frequency or the exact count. In this approach, each document is represented as a binary vector, with each element indicating the presence or absence of a specific term. This method can be useful in scenarios where the frequency information is less relevant, and only the presence of specific terms is significant for classification. It simplifies the representation and reduces the dimensionality of the feature space.

Analysis of the classifiers

C. Naive Bayes Classifier

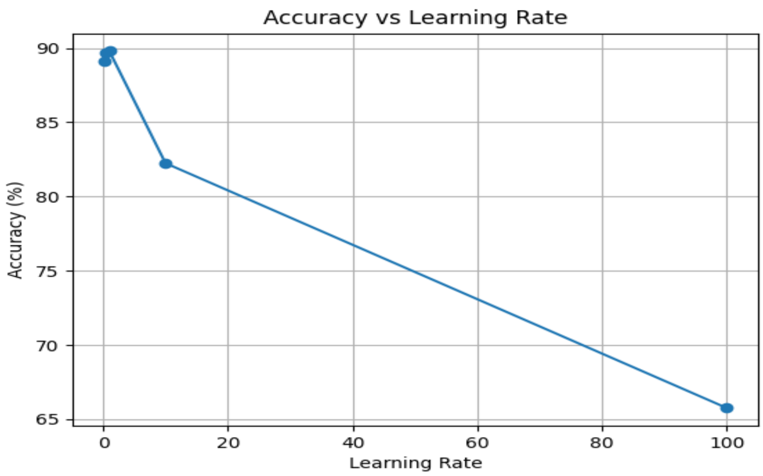
Here are the results for the average accuracy against the different smoothing factors

Smoothing Factor	Bag of Words	TF-IDF	Term Presence
0.1	89.1236	66.7416	84.764
0.5	89.6629	66.427	83.2809
1	89.7978	66.382	83.1011
10	82.2472	66.382	82.7865
100	65.7528	66.382	82.7416

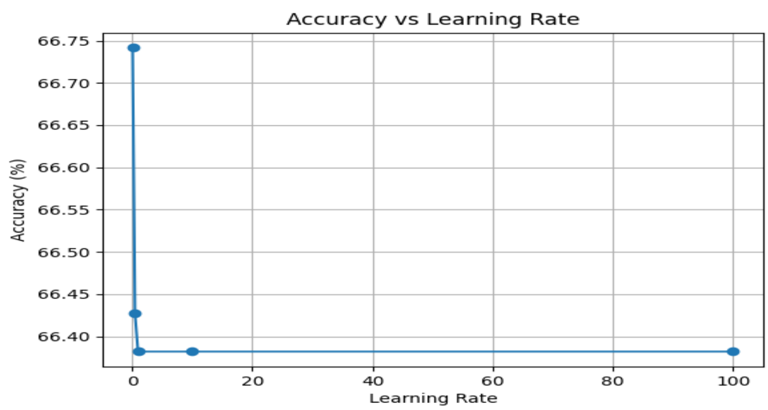
Fig. Average Accuracy for each feature extraction method with a different smoothing factor

Plot Graph

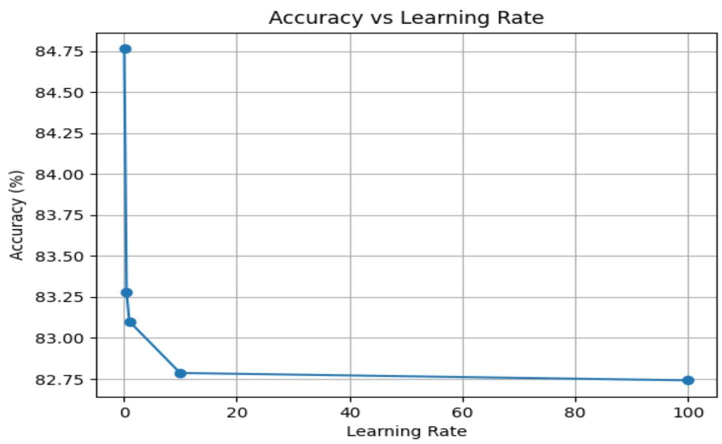
A1. Bag of Words



A2 TD-IDF



A3 Term Presence



D. Logistic Regression

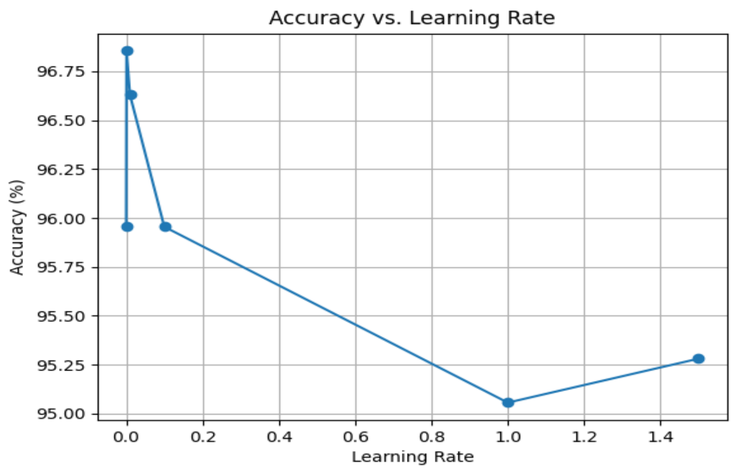
Here are the results for the average accuracy against the different learining rate

Learning Rate	Bag of Words	TF-IDF	Term Presence
0.0001	95.9551	96.6292	96.1798
0.001	96.8539	96.1798	96.4045
0.01	96.6292	95.7303	96.4045
0.1	95.9551	93.9326	95.2809
1	95.0562	95.0562	95.7303
1.5	95.2809	96.1798	95.2809

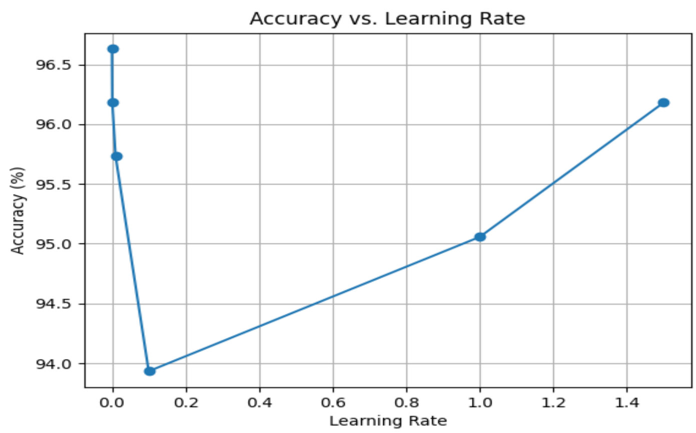
fig. Average accuracy for each feature with a different learning rate

Plot Graph

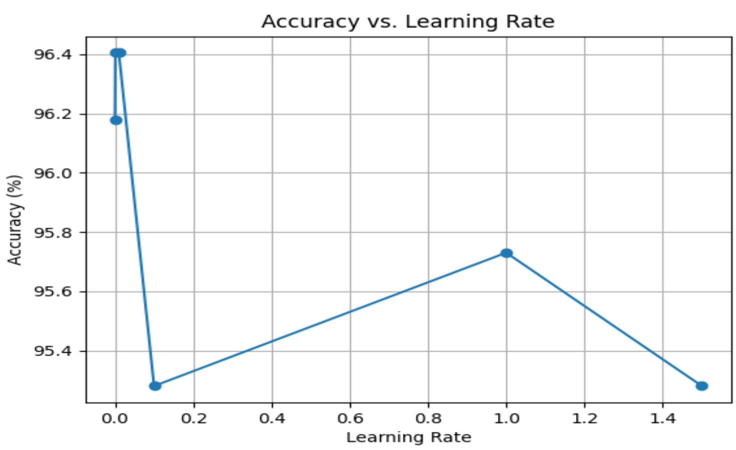
B1. Bag of Words



B2. TD-IDF



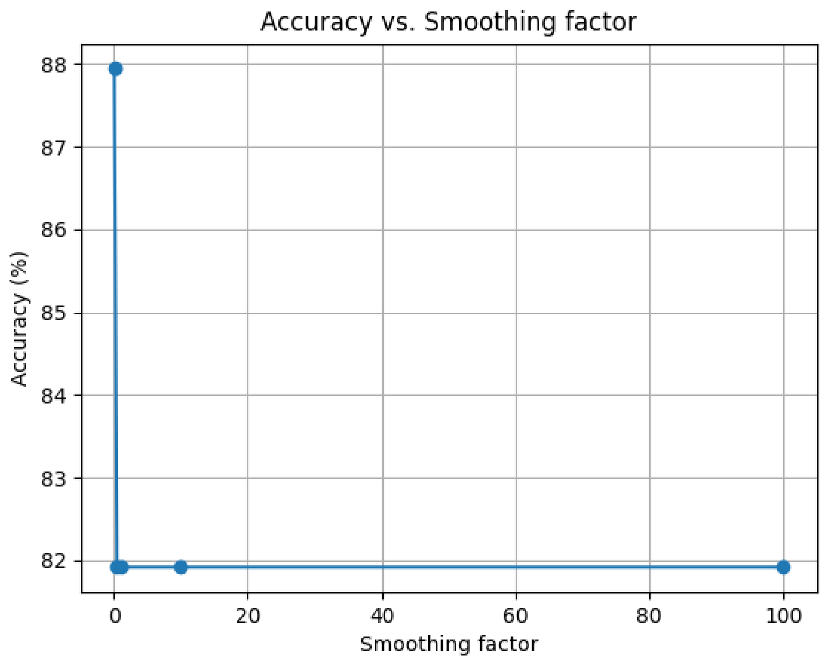
B3 Term presence



3.The Weather Data set

These are the results we found after training the Naive Bays and the Logistic regression models with a weather data that contains fields for various weather elements and gives a class of either yes or no for whether there will be rain tomorrow.

Naive Bayes Classifier



Logistic Regression