//make sure to peruse thru the README and requirements file

The kernel function is very important when it comes to supporting vector machine classification. However, it can be hard to choose the best kernel based on the available alternatives due to the complexity of the problem. Although kernel methods have been widely used in machine learning, identifying the best ones among the given kernels is important to their success. This project presents a comparative evaluation of three kernel functions of the SVM algorithm; radial basis function, sigmoid function, and polynomial function. Evaluation of performance is done to ascertain their general performance. The experiments have been conducted on a multi-classification problem.

**Design Method**

The study used a dataset of scrapped articles from various blogs and websites. The dataset has 7,600 rows and each row holds random data that was captured from the websites' content. To ensure that our evaluation results are comparable to those published in the literature, we focused on the topic of text categorization. The articles' contents were categorized into one of four categories: business, sports, world news, and SciTech . Features were then created by text processing.

The computation and implementation of the algorithms were done in Python using the libraries numpy and scikit-learn. The algorithms were modeled using the same framework to avoid bias in performance estimation and parameter tuning using the following steps;

1. Text processing was done to establish the frequency of words.
2. Data labeling Text labels were turned into numbers for the classifier to ensure creating a good labeled training set with correct labels.
3. There was need to create features from the text. In order to do this, words were turned into numbers for easier analysis. This way weight is given which is a good indicator of how important a word is to a document.
4. Separate a validation dataset in order to establish whether the model created is good. In this study the dataset is split into two sets; 30% training set and 70% for validation.
5. model selection and evaluation is done through exploratory data analysis. The study employs use of gamma and cost function parameters in all four machine learning algorithms.
6. In order to evaluate the learning algorithms, a test harness was set up using the following metrics;
   1. Accuracy - the fraction of samples predicted correctly
   2. Precision - the ratio of true positives to false positives or the ability of the classifier not to label a positive sample as negative
   3. Recall - the ratio of true positives to false negatives or the ability of the classifier to find all the positive samples
   4. F1 Score - The harmonic average of precision and recall
   5. Time taken – computation time taken by the learning algorithm

**Test**

The first experiment implements SVM with an “auto” function. This function allows the model to choose the best SVM kernel to use for the problem at hand without revealing. Tuning parameters value for machine learning algorithms effectively improves the model performance. In the next instance SVM auto is tuned with “gamma” and “c”. Gamma is a kernel coefficient for ‘rbf’, ‘poly’ and ‘sigmoid’. The higher the value of gamma, the more the results will try to exact fit as per the training data set therefore causing a generalization error and cause over-fitting problem.

**Radial Basis Function** can be implemented using;

The adjustable parameter sigma plays a major role in the performance of the kernel, and should be carefully tuned to the problem at hand. If overestimated, the exponential will behave almost linearly and the higher-dimensional projection will start to lose its non-linear power. On the other hand, if underestimated, the function will lack regularization and the decision boundary will be highly sensitive to noise in training data.

The **Polynomial kernel** is a non-stationary kernel. Polynomial kernels are well suited for problems where all the training data is normalized.

The **sigmoid kernel** comes from Neural Networks field, where the bipolar sigmoid function is often used as an activation function for artificial neurons [2].

where d, and are the adjustable parameters of the above kernel functions.

There are two adjustable parameters in the sigmoid kernel, the slope alpha and the intercept constant **c**. A common value for alpha is 1/N, where N is the data dimension.