Objective - The main objective of this project is to build a text classifier using CNN which takes text as the input and predicts the label.

Dataset details – The dataset contains 20,000 (Approx) newsgroup documents from 20 different newsgroups. The dataset is of size 46 MB approximately.

- Data source - *https://archive.ics.uci.edu/dataset/113/twenty+newsgroups (https://archive.ics.uci.edu/dataset/113/twenty+newsgroups)*
- Data size (MB, GB, etc.) - 46 MB approx.
- Data shape (# of rows and # columns) - It does not contain data in the form of rows and columns rather data is present in the form text files (.txt format)

Research Question – Convolution Neural Networks are usually used for image processing. So, for this study I am going to analyze if it can also be used for processing text.

Approach

- The data here contains in the form of text files and does not contain any tabular format data.
- The text documents are first cleaned and processed using python regex which involves removing punctuations, converting non-English words to English etc.,
- The cleaned text documents are then vectorized.
- We use pre-trained glove vectors for vectorizing the text.
- CNN architectures are built and are trained using the above vectorized documents.
- Results are evaluated for the train, test datasets.

```
In [1]:   1  # Spacy is a natural processing language library in Python which provides
          2  # various languages.
          3  # Installing Spacy
          4  !pip install spacy
```

```
Requirement already satisfied: spacy in /usr/local/lib/python3.10/dist-packag
es (3.6.1)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/
python3.10/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/
python3.10/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/py
thon3.10/dist-packages (from spacy) (1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.
10/dist-packages (from spacy) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python
3.10/dist-packages (from spacy) (3.0.9)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /usr/local/lib/python3.
10/dist-packages (from spacy) (8.1.12)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python
3.10/dist-packages (from spacy) (1.1.2)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.
10/dist-packages (from spacy) (2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/pyth
on3.10/dist-packages (from spacy) (2.0.10)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /usr/local/lib/python
3.10/dist-packages (from spacy) (0.9.0)
Requirement already satisfied: pathy>=0.10.0 in /usr/local/lib/python3.10/dis
t-packages (from spacy) (0.10.3)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/pyt
hon3.10/dist-packages (from spacy) (6.4.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.
10/dist-packages (from spacy) (4.66.1)
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.10/dis
t-packages (from spacy) (1.23.5)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/pyth
on3.10/dist-packages (from spacy) (2.31.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/l
ocal/lib/python3.10/dist-packages (from spacy) (1.10.13)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packa
ges (from spacy) (3.1.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-p
ackages (from spacy) (67.7.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/d
ist-packages (from spacy) (23.2)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/pyth
on3.10/dist-packages (from spacy) (3.3.0)
Requirement already satisfied: typing-extensions>=4.2.0 in /usr/local/lib/pyt
hon3.10/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy) (4.
5.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/pyt
hon3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist
-packages (from requests<3.0.0,>=2.13.0->spacy) (3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.1
0/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.1
0/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2023.11.17)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.1
0/dist-packages (from thinc<8.2.0,>=8.1.8->spacy) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/pyt
hon3.10/dist-packages (from thinc<8.2.0,>=8.1.8->spacy) (0.1.4)
```

```
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.
10/dist-packages (from typer<0.10.0,>=0.3.0->spacy) (8.1.7)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/d
ist-packages (from jinja2->spacy) (2.1.3)
```

In [2]:
```python
# Installing english language for spacy

!python -m spacy download en
```

```
2023-12-15 04:16:02.477913: E external/local_xla/xla/stream_executor/cuda/cud
a_dnn.cc:9261] Unable to register cuDNN factory: Attempting to register facto
ry for plugin cuDNN when one has already been registered
2023-12-15 04:16:02.477970: E external/local_xla/xla/stream_executor/cuda/cud
a_fft.cc:607] Unable to register cuFFT factory: Attempting to register factor
y for plugin cuFFT when one has already been registered
2023-12-15 04:16:02.479883: E external/local_xla/xla/stream_executor/cuda/cud
a_blas.cc:1515] Unable to register cuBLAS factory: Attempting to register fac
tory for plugin cuBLAS when one has already been registered
2023-12-15 04:16:02.491176: I tensorflow/core/platform/cpu_feature_guard.cc:1
82] This TensorFlow binary is optimized to use available CPU instructions in
performance-critical operations.
To enable the following instructions: AVX2 FMA, in other operations, rebuild
TensorFlow with the appropriate compiler flags.
2023-12-15 04:16:04.188699: W tensorflow/compiler/tf2tensorrt/utils/py_utils.
cc:38] TF-TRT Warning: Could not find TensorRT
2023-12-15 04:16:07.067587: I external/local_xla/xla/stream_executor/cuda/cud
a_executor.cc:901] successful NUMA node read from SysFS had negative value (-
1), but there must be at least one NUMA node, so returning NUMA node zero. Se
e more at https://github.com/torvalds/linux/blob/v6.0/Documentation/ABI/testi
ng/sysfs-bus-pci#L344-L355 (https://github.com/torvalds/linux/blob/v6.0/Docum
entation/ABI/testing/sysfs-bus-pci#L344-L355)
2023-12-15 04:16:07.068201: I external/local_xla/xla/stream_executor/cuda/cud
a_executor.cc:901] successful NUMA node read from SysFS had negative value (-
1), but there must be at least one NUMA node, so returning NUMA node zero. Se
e more at https://github.com/torvalds/linux/blob/v6.0/Documentation/ABI/testi
ng/sysfs-bus-pci#L344-L355 (https://github.com/torvalds/linux/blob/v6.0/Docum
entation/ABI/testing/sysfs-bus-pci#L344-L355)
2023-12-15 04:16:07.068482: I external/local_xla/xla/stream_executor/cuda/cud
a_executor.cc:901] successful NUMA node read from SysFS had negative value (-
1), but there must be at least one NUMA node, so returning NUMA node zero. Se
e more at https://github.com/torvalds/linux/blob/v6.0/Documentation/ABI/testi
ng/sysfs-bus-pci#L344-L355 (https://github.com/torvalds/linux/blob/v6.0/Docum
entation/ABI/testing/sysfs-bus-pci#L344-L355)
⚠ As of spaCy v3.0, shortcuts like 'en' are deprecated. Please use the
full pipeline package name 'en_core_web_sm' instead.
Collecting en-core-web-sm==3.6.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_
core_web_sm-3.6.0/en_core_web_sm-3.6.0-py3-none-any.whl (https://github.com/e
xplosion/spacy-models/releases/download/en_core_web_sm-3.6.0/en_core_web_sm-
3.6.0-py3-none-any.whl) (12.8 MB)
                                      ──────────────────────────── 12.8/12.8 MB 21.4 MB/s eta 0:0
0:00
Requirement already satisfied: spacy<3.7.0,>=3.6.0 in /usr/local/lib/python3.
10/dist-packages (from en-core-web-sm==3.6.0) (3.6.1)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/
python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0)
(3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/
python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0)
(1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/py
thon3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (1.
0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.
10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python
```

```
3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (3.0.9)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /usr/local/lib/python3.
10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (8.1.12)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python
3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (1.1.2)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.
10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/pyth
on3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.0.1
0)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /usr/local/lib/python
3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (0.9.0)
Requirement already satisfied: pathy>=0.10.0 in /usr/local/lib/python3.10/dis
t-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (0.10.3)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/pyt
hon3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (6.4.
0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.
10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (4.66.1)
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.10/dis
t-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (1.23.5)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/pyth
on3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.31.
0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/l
ocal/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==
3.6.0) (1.10.13)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packa
ges (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (3.1.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-p
ackages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (67.7.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/d
ist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (23.2)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/pyth
on3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (3.3.
0)
Requirement already satisfied: typing-extensions>=4.2.0 in /usr/local/lib/pyt
hon3.10/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy<3.7.
0,>=3.6.0->en-core-web-sm==3.6.0) (4.5.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/pyt
hon3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.7.0,>=3.6.0->en-
core-web-sm==3.6.0) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist
-packages (from requests<3.0.0,>=2.13.0->spacy<3.7.0,>=3.6.0->en-core-web-sm=
=3.6.0) (3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.1
0/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.7.0,>=3.6.0->en-core-w
eb-sm==3.6.0) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.1
0/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.7.0,>=3.6.0->en-core-w
eb-sm==3.6.0) (2023.11.17)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.1
0/dist-packages (from thinc<8.2.0,>=8.1.8->spacy<3.7.0,>=3.6.0->en-core-web-s
m==3.6.0) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/pyt
hon3.10/dist-packages (from thinc<8.2.0,>=8.1.8->spacy<3.7.0,>=3.6.0->en-core
-web-sm==3.6.0) (0.1.4)
```

```
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.
10/dist-packages (from typer<0.10.0,>=0.3.0->spacy<3.7.0,>=3.6.0->en-core-web
-sm==3.6.0) (8.1.7)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/d
ist-packages (from jinja2->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.1.
3)
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
```

In [3]:
```python
1  import zipfile
2  import re
3  import spacy
4  from tqdm import tqdm
```

In [4]:
```python
1  from google.colab import drive
2  drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

In [5]:
```python
1  #We have dataset in the form of zip file
2  #Unzipping the zip file.
3
4  z = zipfile.ZipFile('/content/drive/MyDrive/Srisailam/20_newsgroups.zip')
5  z.extractall()
```

In [6]:
```python
1   #extracting the class labels from the file paths to check its value counts
2   import os
3
4   dir_labels = []
5   dir_paths  = []
6   for root, dirs, files in os.walk("20_newsgroups", topdown=False):
7       for name in files:
8           x = os.path.join(root, name)
9           dir_labels.append(x.split('/')[-2])
10          dir_paths.append(x)
```

In [7]:
```python
1  len(dir_labels)
```

Out[7]: 19997

There are 19997 documents in the dataset.

```
In [8]:   1  import numpy as np
          2  list(np.unique(np.array(dir_labels)))
```

```
Out[8]:  ['alt.atheism',
          'comp.graphics',
          'comp.os.ms-windows.misc',
          'comp.sys.ibm.pc.hardware',
          'comp.sys.mac.hardware',
          'comp.windows.x',
          'misc.forsale',
          'rec.autos',
          'rec.motorcycles',
          'rec.sport.baseball',
          'rec.sport.hockey',
          'sci.crypt',
          'sci.electronics',
          'sci.med',
          'sci.space',
          'soc.religion.christian',
          'talk.politics.guns',
          'talk.politics.mideast',
          'talk.politics.misc',
          'talk.religion.misc']
```

There are a total of 20 unique class labels as shown above.

Exploratory Data Analysis:

Figure 1:

```
In [9]:   1  import matplotlib.pyplot as plt
          2  import pandas as pd
```

```
In [10]:    1  #Bar chart showing the counts of each class label
            2
            3  plt.figure(figsize=(12, 8))
            4  plt.xticks(fontsize=12)
            5  plt.yticks(fontsize=12)
            6  ax = pd.Series(dir_labels).value_counts().sort_index().plot(kind='bar',col
            7
            8  for c in ax.containers:
            9      # Optional: if the segment is small or 0, customize the labels
           10      labels = [v.get_height() if v.get_height() > 0 else '' for v in c]
           11      # remove the labels parameter if it's not needed for customized labels
           12      ax.bar_label(c, labels=labels, label_type='edge')
           13
           14  plt.xticks(rotation = 90)
           15  plt.title("Bar Plot of the Class labels",fontweight='bold')
           16  plt.ylabel("Counts",fontsize=12,fontweight='bold')
           17  plt.xlabel("Class labels",fontsize=12,fontweight='bold')
           18  plt.show()
```

Bar Plot of the Class labels
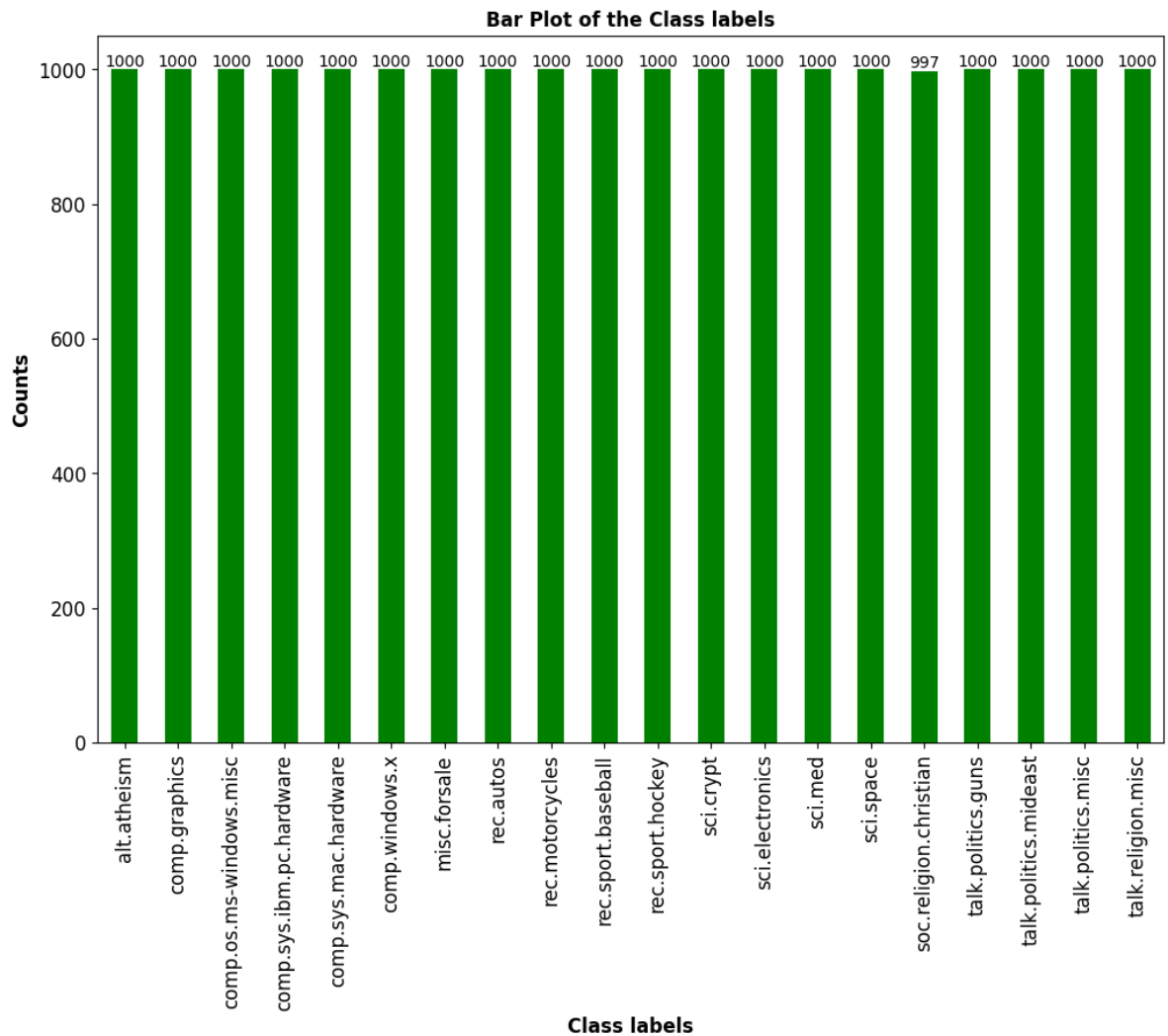
Observations:

From the above bar chart it is clear that

- All the class labels has equal number of documents except for the soc.religion.christian class label.
- soc.religion.christian class label has 997 documents while all the other class labels have 1000 documents.

Figure 2:

```
In [11]:
1  # Line chart showing the mean file size of each newsgroup.
2
3  file_size = {}
4  for root, dirs, files in os.walk("20_newsgroups", topdown=False):
5      for name in files:
6          x = os.path.join(root, name)
7          group_label = x.split('/')[-2]
8          if group_label in file_size:
9              file_size[group_label].append(os.path.getsize(x))
10          else:
11              file_size[group_label] = [os.path.getsize(x)]
```

```
In [12]:
1  file_sizes_mean = [np.array(file_size[i]).mean() for i in file_size.keys()
```

```
In [13]:
1  plt.figure(figsize=(12, 8))
2  plt.xticks(fontsize=10)
3  plt.yticks(fontsize=10)
4
5  plt.title("Line chart showing mean file sizes of each newsgroup in KB",for
6  plt.xlabel("Mean file size in KB",fontsize=10,fontweight='bold')
7  plt.ylabel("Newsgroup",fontsize=10,fontweight='bold')
8  ax = plt.plot(file_sizes_mean,list(np.unique(np.array(dir_labels))))
9
10  plt.grid('--')
11  plt.show()
```


Line chart showing mean file sizes of each newsgroup in KB
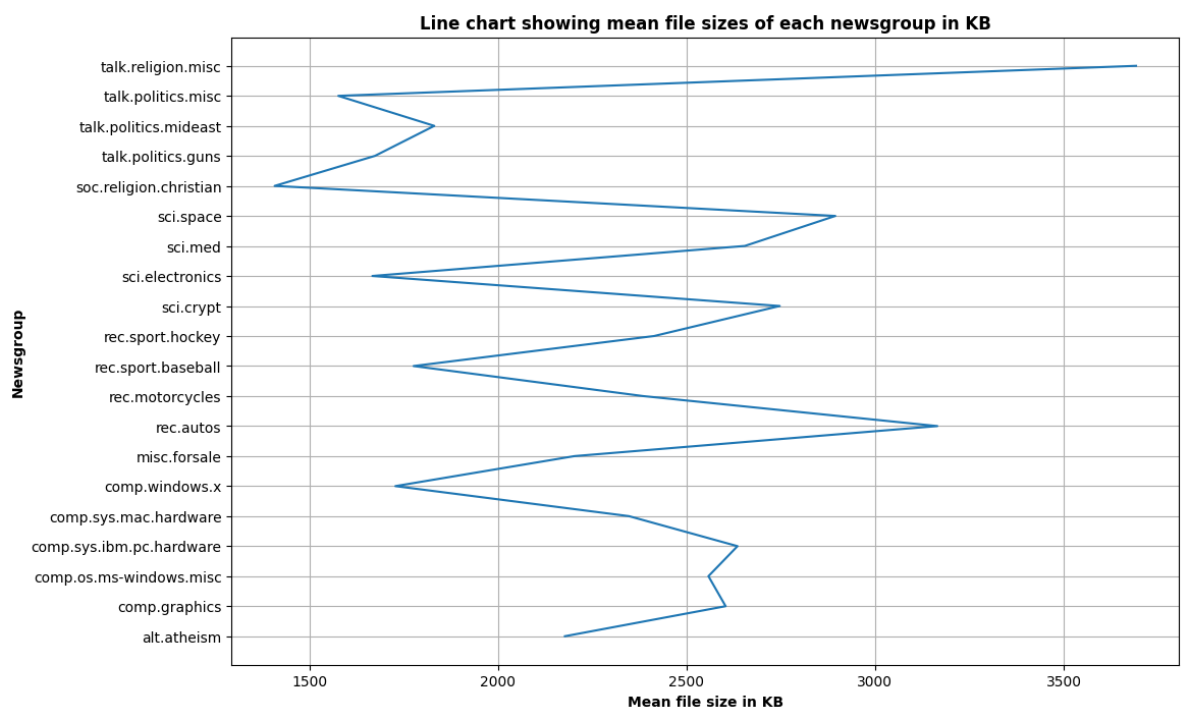
Observations:

From the above line chart it can be observed that the

- talk.religion.misc has got the highest mean file size.
- soc.religion.christian has got the lowest mean file size.

Progress for Presentation - 2

Data Cleaning and Preparation

In [14]:
```python
1  # printing text in the random article.
2
3  print("Name of the article - ",dir_paths[1])
4  print('\n')
5  file = open(dir_paths[1],encoding="utf8", errors='ignore')
6  info = file.read()
7  print(info)
8  file.close()
```

Name of the article - 20_newsgroups/comp.graphics/38937

Path: cantaloupe.srv.cs.cmu.edu!magnesium.club.cc.cmu.edu!pitt.edu!uunet!mdis
ea!mothost!lmpsbbs!davidla@ecs.comm.mot.com
From: davidla@ecs.comm.mot.com (David Lau)
Newsgroups: comp.graphics
Subject: GIF viewer source code for PC?
Message-ID: <1993Apr29.152235.22224@lmpsbbs.comm.mot.com>
Date: 29 Apr 93 15:22:35 GMT
Sender: news@lmpsbbs.comm.mot.com (Net News)
Organization: Motorola
Lines: 10
Nntp-Posting-Host: 145.1.161.169

Does anyone know where I could get some source code for a GIF viewer on a PC.
It doesn't have to be fancy, it would be best if it were written in C to
display gif pictures on a VGA screen. I was hoping to get it for free
from an FTP site, but I haven't been able to find it yet.

David Lau
davidla@ecs.comm.mot.com

In [15]:
```python
def emailprocessing(input_text):
    """
    This function takes the text as the input in
    string format and processes the words with
    email extensions and returns a string which
    is concatenated with list of such words.
    """
    raw_text  = input_text.read()
    emails = re.findall(r"[\w\-\.]+@[\w\.-]+\b", raw_text)# extracts email
    processed_email   = []
    for i in emails:
        i = i.split('@')[1] # split the word based on '@'
        temp = i.split('.')
        for j in temp:
            if len(j)<=2:# check for the length of the word if it is less
                continue
            elif j=='com':# check if the word is 'com' and ignore if its t
                continue
            elif "@" in j:
                processed_email.append(j[1:])
            else:
                processed_email.append(j)
    return ' '.join(processed_email).lower()
```

In [16]:
```python
# Applying emailprocessing function for the above text in the article
# and printing its output

file = open(dir_paths[1],encoding="utf8", errors='ignore')
print(emailprocessing(file))
file.close()
```

ecs comm mot ecs comm mot lmpsbbs comm mot lmpsbbs comm mot ecs comm mot

In [17]:
```python
def subjectprocessing(input_text):
    """
    This function takes the text as the input and
    processes the text line after the word "Subject:"
    and returns the string.
    """
    raw_text  = input_text.read()
    subject = re.findall('^Subject.*$', raw_text,re.MULTILINE) # extracts
                                                 # that star
                                                 # as list o

    processed_subject  = []
    subject = re.sub('[^A-Za-z0-9]+', ' ', subject[0]) # Replace non alpho
                                         # single space.

    subject = subject.rstrip().split(' ')# removes the leading and trailin
                                         # and splits the text into words
    for i in subject:
        if i=='Subject' or i=='Re':# checks if the word is "Subject" or "F
            continue
        else:
            processed_subject.append(i)
    return ' '.join(processed_subject).lower()# joins the list of words in
```

In [18]:
```python
# Applying subjectprocessing function for the above text in the article
# and printing its output

file = open(dir_paths[1],encoding="utf8", errors='ignore')
print(subjectprocessing(file))
file.close()
```

gif viewer source code for pc

In [19]:
```python
def textprocessing(input_text):

    raw_text  = input_text.read()
    text  =  re.sub(r"[\w\-\.]+@[\w\.-]+\b",' ',raw_text) # finds all occu
                                                          # and replaces t
    text  =  re.sub(r'^Subject.*$',' ',text,flags=re.MULTILINE) # finds al
                                                          # that sto
                                                          # a single
    text  =  re.sub(r'Write to:.*$','', text,flags=re.MULTILINE) # replace
                                                          # an empt
    text  =  re.sub(r'From:.*$','', text,flags=re.MULTILINE)     # replace
                                                          # an empt
    #Decontractions
    text  =  re.sub("[a-zA-Z]+:",'',text) # finds all occurrences of seque
                                          # followed by a colon, and repla
    text  =  re.sub(r"won\'t", "will not", text)  # replaces the words wit
    text  =  re.sub(r"can\'t", "can not", text)   # replaces the words wit
    text  =  re.sub(r"n\'t", " not", text)        # replaces the words wit
    text  =  re.sub(r"\'re", " are", text)        # replaces the words wit
    text  =  re.sub(r"\'s", " is", text)          # replaces the words wit
    text  =  re.sub(r"\'d", " would", text)       # replaces the words wit
    text  =  re.sub(r"\'ll", " will", text)       # replaces the words wit
    text  =  re.sub(r"\'t", " not", text)         # replaces the words wit
    text  =  re.sub(r"\'ve", " have", text)       # replaces the words wit
    text  =  re.sub(r"\'m", " am", text)          # replaces the words wit
    text  =  text.strip()                         # remove the Leading and
    text  =  re.sub(r'\<[^<>]*\>','', text,flags=re.MULTILINE) # removes t
                                                          # and does
    text  =  re.sub(r"\([^()]*\)",'', text,flags=re.MULTILINE) # removes t
                                                          # and does
    text  =  re.sub(r'\[[^[]*\]','', text,flags=re.MULTILINE)  # removes t
                                                          # and does
    text  =  text.replace('>','')  # replaces the parenthesis > with an em
    text  =  text.replace('\\','') # replaces the backslash with an empty
    text  =  text.replace('?',' ') # replaces the '?' with an empty string
    text  =  re.sub('\n',' ', text,flags=re.MULTILINE) # replaces all the
                                                          # and does this for
    text  =  re.sub(' +',' ',text,flags=re.MULTILINE)  # replaces all the
                                                          # and does this for
    text  =  re.sub('-',' ', text,flags=re.MULTILINE)  # replaces multiple
                                                          # and does this for
    # Spacy is a natural processing language library in Python which provi
    # various languages including english.
    NER = spacy.load("en_core_web_sm") # loads the english language model.
    text1= NER(text) # text is tokenized and further linguistic analyses c
    a = []
    for i in text1.ents: # loops through the named entities (e.g., names o
                         # from the tokens.
        a.append((i.text,i.label_)) # appends the text of recognized named
    for i in a:
        if i[1]=='PERSON': # checks if the label of the named entity is a
            text = text.replace(i[0],'') # replace person entities in the
        elif i[1]=='GPE': # checks if the label of the named entity is geo
            if " " in i[0]:
                temp = i[0].replace(' ','_')
                text = text.replace(i[0],temp) # if the word is a geopolit
                                                          # with an underscore to con
```

```
58        text = re.sub('\d',' ',text) #replaces digits(0-9) with a single space
59        text = re.sub(r'\b_+|_+\b', '', text) # finds the underscores that app
60                                    # or at the end or have one or r
61                                    # and replaces with an empty str
62        text = re.sub(r"\b[a-zA-Z]{1}_([a-zA-Z]+)",r"\1",text) #  It searches
63                                         #  a single let
64                                         #  followed by
65                                         # it replaces i
66                                         # the one or mc
67        text = re.sub(r"\b[a-zA-Z]{2}_([a-zA-Z]+)",r"\1",text) # It searches j
68                                         # two alphabeti
69                                         # underscore, c
70                                         # it replaces i
71                                         # the one or mc
72        text = text.lower()
73        text = re.sub(r'\b\w{1,2}\b',' ',text) # It searches for words (sequer
74                                         # that consist of one or two ch
75        text = re.sub(r'\b\w{15,}\b',' ',text) # It searches for words (sequer
76                                         # that consist of 15 or more ch
77        text = re.sub(r'[^a-zA-Z_]',' ',text)  # It searches for characters th
78                                         # (both uppercase and lowercase
79                                         # them with space characters.
80        text = re.sub(r' {2,}',' ',text)     # It searches for two or more c
81                                         # and replaces them with a sing
82        return text
```

In [20]:
```
1  # Applying textprocessing function for the above text in the article
2  # and printing its output
3
4  file = open(dir_paths[1],encoding="utf8", errors='ignore')
5  print(textprocessing(file))
6  file.close()
```

cantaloupe srv cmu edu magnesium club cmu edu pitt edu uunet mdisea mothost l
mpsbbs comp graphics message apr gmt motorola nntp posting does anyone know w
here could get some source code for gif viewer does not have fancy would best
were written display pictures vga screen was hoping get for free from ftp sit
e but have not been able find yet

In [21]:
```
1  email = []
2  for i in tqdm(dir_paths): # Looping through all the files and
3                            # applying email processing function and
4                            # appending the output of the funtion to the
5                            # email list.
6      sample = open(i,encoding="utf8", errors='ignore')
7      email.append(emailprocessing(sample))
8      sample.close()
```

100%|████████| 19997/19997 [00:08<00:00, 2333.04it/s]

In [22]:
```python
subject = []
for i in tqdm(dir_paths): # looping through all the files and
                          # applying subject processing function and
                          # appending the output of the funtion to the
                          # subject list.
    sample = open(i,encoding="utf8", errors='ignore')
    subject.append(subjectprocessing(sample))
    sample.close()
```

100%|████████████| 19997/19997 [00:02<00:00, 8235.66it/s]

In [ ]:
```python
text = []
for i in tqdm(dir_paths): # looping through all the files and
                          # applying text processing function and
                          # appending the output of the funtion to the
                          # text list.
    sample = open(i,encoding="utf8", errors='ignore')
    text.append(textprocessing(sample))
    sample.close()
```

In [ ]:
```python
#creating a dataframe out of email, subject, text and dir_labels lists

data = pd.DataFrame(list(zip(email,subject,text,dir_labels)),columns = ['e
data.to_pickle("processed_data.pkl") # storing the dataframe in a pickle f
```

In [23]:
```python
unpickled_df = pd.read_pickle('/content/drive/MyDrive/Srisailam/processed_
unpickled_df.head()
```

Out[23]:

	email	subject	text	target_label
0	mantis mantis mit edu mantis netcom mantis	alt atheism faq atheist resources	cantaloupe srv cmu edu alt alt atheism news al...	alt.atheism
1	mantis mantis mit edu mantis mantis mantis	alt atheism faq introduction to atheism	cantaloupe srv cmu edu alt alt atheism news al...	alt.atheism
2	dbstu1 tu-bs dbstu1 tu-bs ibr tu-bs dbstu1 tu-...	gospel dating	alt atheism cantaloupe srv cmu edu crabapple s...	alt.atheism
3	mantis mantis kepler unh edu kepler unh edu	university violating separation of church state	cantaloupe srv cmu edu alt alt politics usa ca...	alt.atheism
4	watson ibm com watson ibm austin ibm cbnewsl a...	soc motss et al princeton axes matching funds ...	cantaloupe srv cmu edu alt soc rec alt atheism...	alt.atheism

In [24]:
```python
# extracting the length of the words for each row in the text
# and assigning to new column by name text_word_count.

unpickled_df['text_word_count'] = unpickled_df['text'].apply(lambda x:len(
```

In [25]:
```
1  # printing first 5 rows of the data.
2  unpickled_df.head()
```

Out[25]:

	email	subject	text	target_label	text_word_count
0	mantis mantis mit edu mantis netcom mantis	alt atheism faq atheist resources	cantaloupe srv cmu edu alt alt atheism news al...	alt.atheism	1224
1	mantis mantis mit edu mantis mantis mantis	alt atheism faq introduction to atheism	cantaloupe srv cmu edu alt alt atheism news al...	alt.atheism	4023
2	dbstu1 tu-bs dbstu1 tu-bs ibr tu-bs dbstu1 tu-...	gospel dating	alt atheism cantaloupe srv cmu edu crabapple s...	alt.atheism	518
3	mantis mantis kepler unh edu kepler unh edu	university violating separation of church state	cantaloupe srv cmu edu alt alt politics usa ca...	alt.atheism	189
4	watson ibm com watson ibm austin ibm cbnewsl a...	soc motss et al princeton axes matching funds ...	cantaloupe srv cmu edu alt soc rec alt atheism...	alt.atheism	107

In [26]:
```
1  plt.figure(figsize=(10,5))
2  unpickled_df['text_word_count'].plot.hist(color='green')
3
4  plt.title("Distribution of word counts in text column",fontsize=8,fontweig
5  plt.xlabel("Word counts",fontsize=8,fontweight='bold')
6  plt.ylabel("Frequency",fontsize=8,fontweight='bold')
7  plt.show()
```


Distribution of word counts in text column
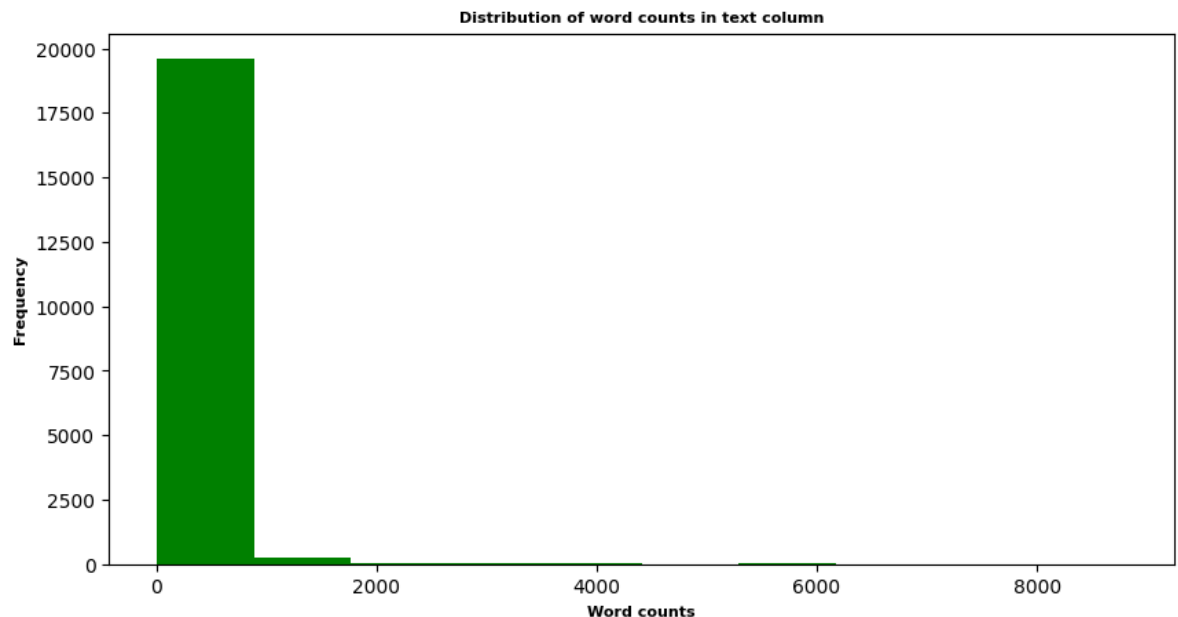
Observation:
From the above chart, it can be observed that the most of the word counts in the text column are less than 1000.

Displaying Wordclouds for subject and text columns

In [27]:
```python
from wordcloud import WordCloud
```

In [28]:
```python
#Word Cloud for text column

wordcloud = WordCloud(width=800, height=400, background_color="white").\
            generate(" ".join(unpickled_df['text'].tolist()))
# Display the word cloud using matplotlib
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.title("Word Cloud for text column",fontsize=9,fontweight='bold')
plt.axis("off")
plt.show()
```


Word Cloud for text column

Observations:

It can be observed that the Ohio state is highly used word in the news article. News articles must have covered most of its news from Ohio state.

In [29]:
```python
# word cloud for subject column

wordcloud = WordCloud(width=800, height=400, background_color="white").\
            generate(" ".join(unpickled_df['subject'].tolist()))
# Display the word cloud using matplotlib
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.title("Word Cloud for subject column",fontsize=9,fontweight='bold')
plt.axis("off")
plt.show()
```


Observations:

It can be observed that the sale and year are highly used words in the subject of news articles.

Train-Test Split

In [30]:
```python
# combining the Processed_email , Processed_subject , Processed_text colum
import numpy as np

X = np.array(unpickled_df['email']+unpickled_df['subject']+unpickled_df['t
```

In [31]:
```python
y = np.array(unpickled_df['target_label'])
```

In [32]:
```python
# Splitting the data into train and test with test size of 25% of the tota

from sklearn.model_selection import train_test_split
from sklearn import preprocessing
import tensorflow as tf

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25,
le = preprocessing.LabelEncoder() # Encoding the target labels using label
le.fit(y_train)
y_train  = le.transform(y_train)
y_test   = le.transform(y_test)
y_train  = tf.keras.utils.to_categorical(y_train) # converting labels in y
y_test   = tf.keras.utils.to_categorical(y_test)  # converting labels in y
```

In [33]:
```python
# printing the shapes of train and test datasets

print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(14997,)
(5000,)
(14997, 20)
(5000, 20)
```

In [34]:
```python
lengths = [len(s.split()) for s in X_train] # finding length of words in e
max_length = int(np.percentile(lengths,99)) # finding the length pertainin
print(max_length)
```

```
1361
```

The above max length gives information that 99% of the documents have less than or equal 1361 words in it.

Modelling

In [35]:
```python
import tensorflow
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences

tokenizer = Tokenizer(num_words=10000, lower= 1,\
                      filters= '!"#$%&()*+,-./:;<=>?@[\\]^`{|}~\t\n',oov_t
                                                  # the text data excl
                                                  # in the filters as
tokenizer.fit_on_texts(X_train)
train_sequences = tokenizer.texts_to_sequences(X_train)
test_sequences  = tokenizer.texts_to_sequences(X_test)

# The pad_sequences function pads or truncates the input sequences to have
# and it adds padding at the end of each sequence. The result is a set of

train_padded = pad_sequences(train_sequences,padding="post",\
                             truncating='post',maxlen=int(0.1*max_length),
test_padded  = pad_sequences(test_sequences,padding="post",\
                             truncating='post',maxlen=int(0.1*max_length),

# printing the shape of train and test datasets post padding the sequences
print(train_padded.shape)
print(test_padded.shape)
```

```
(14997, 136)
(5000, 136)
```

In [36]:
```python
size_of_vocabulary = len(tokenizer.word_index) + 1
print(size_of_vocabulary)
```

```
94450
```

In [38]:
```python
#Reference - https://www.analyticsvidhya.com/blog/2020/03/pretrained-word-

# creating a dictionary(word:coeffs) using glove vectors text file.
embeddings_index = dict()
f = open('/content/drive/MyDrive/Srisailam/glove.6B.100d.txt',encoding="ut
for line in f:
    values = line.split()
    word   = values[0]
    coefs  = np.asarray(values[1:],dtype='float32')
    embeddings_index[word] = coefs
f.close()
print('Loaded {0} word vectors'.format(len(embeddings_index)))
```

```
Loaded 400000 word vectors
```

```python
In [39]:  1  # finding the embedding matrix for the words in the tokenizer using
          2  # embedding index created earlier.
          3
          4  embedding_matrix = np.zeros((size_of_vocabulary,100))
          5  for word, i in tokenizer.word_index.items():
          6      embedding_vector = embeddings_index.get(word)
          7      if embedding_vector is not None:
          8          embedding_matrix[i] = embedding_vector
```

CNN Architecture

```python
In [64]:  1  from keras.layers import Dense,Input,Activation,Embedding,Conv1D,concatena
          2                     MaxPooling1D,Flatten,Dropout,BatchNormalization
          3  from keras.models import Model
          4  from keras.utils  import plot_model
          5
          6  initializer = tf.keras.initializers.GlorotNormal(seed=0)
          7  input_layer = Input(shape = (int(0.1*max_length),),dtype='int32')
          8  word_embedding = Embedding(size_of_vocabulary,100,weights=[embedding_matri
          9                     input_length=1000,trainable=True)(input_layer)
         10  conv1    =    Conv1D(filters=32,kernel_size=4,activation='relu',padding='sa
         11                    kernel_initializer=initializer)(word_embedding)
         12  conv2    =    Conv1D(filters=32,kernel_size=6,activation='relu',padding='sa
         13                    kernel_initializer=initializer)(word_embedding)
         14  conv3    =    Conv1D(filters=32,kernel_size=8,activation='relu',padding='sa
         15                    kernel_initializer=initializer)(word_embedding)
         16  merged1  =    concatenate([conv1,conv2,conv3])
         17  pool1    =    MaxPooling1D(pool_size=2,padding='same')(merged1)
         18  conv4    =    Conv1D(filters=32,kernel_size=4,activation='relu',padding='sa
         19  conv5    =    Conv1D(filters=32,kernel_size=6,activation='relu',padding='sa
         20  conv6    =    Conv1D(filters=32,kernel_size=8,activation='relu',padding='sa
         21  merged2  =    concatenate([conv4,conv5,conv6])
         22  pool2    =    MaxPooling1D(pool_size=2,padding='same')(merged2)
         23  conv7    =    Conv1D(filters=32,kernel_size=4,activation='relu',padding='sa
         24  flat1    =    Flatten()(conv7)
         25  drop1    =    Dropout(0.5)(flat1)
         26  dense1   =    Dense(50,activation='relu',kernel_initializer=initializer)(dr
         27  output   =    Dense(20,activation='softmax',kernel_initializer=initializer)
         28
         29  model    =    Model(inputs = input_layer,outputs= output)
```

```python
In [65]:  1  from keras.optimizers import Adam
          2  adam = Adam(learning_rate=0.0001)
          3  model.compile(loss='categorical_crossentropy',optimizer = adam ,metrics=[
```

In [66]:
```
1  plot_model(model,show_shapes=True,show_trainable=True)
```

Out[66]:

	input_2	input:	[(None, 136)]
T	InputLayer	output:	[(None, 136)]

	embedding_1	input:	(None, 136)
T	Embedding	output:	(None, 136, 100)

	conv1d_7	input:	(None, 136, 100)
T	Conv1D	output:	(None, 136, 32)

	conv1d_8	input:	(None, 136, 100)
T	Conv1D	output:	(None, 136, 32)

	conv1d_9	input:	(None, 136, 100)
T	Conv1D	output:	(None, 136, 32)

	concatenate_2	input:	[(None, 136, 32), (None, 136, 32), (None, 136, 32)]
T	Concatenate	output:	(None, 136, 96)

	max_pooling1d_2	input:	(None, 136, 96)
T	MaxPooling1D	output:	(None, 68, 96)

	conv1d_10	input:	(None, 68, 96)
T	Conv1D	output:	(None, 68, 32)

	conv1d_11	input:	(None, 68, 96)
T	Conv1D	output:	(None, 68, 32)

	conv1d_12	input:	(None, 68, 96)
T	Conv1D	output:	(None, 68, 32)

	concatenate_3	input:	[(None, 68, 32), (None, 68, 32), (None, 68, 32)]
T	Concatenate	output:	(None, 68, 96)

	max_pooling1d_3	input:	(None, 68, 96)
T	MaxPooling1D	output:	(None, 34, 96)

	conv1d_13	input:	(None, 34, 96)
T	Conv1D	output:	(None, 34, 32)

	flatten_1	input:	(None, 34, 32)
T	Flatten	output:	(None, 1088)

	dropout_1	input:	(None, 1088)
T	Dropout	output:	(None, 1088)

	dense_2	input:	(None, 1088)
T	Dense	output:	(None, 50)

	dense_3	input:	(None, 50)
T	Dense	output:	(None, 20)

In [67]:
```python
1  model.summary(show_trainable=True,positions=[0.35,0.6,0.7,1])
```

Model: "model_1"

_____
_____

Layer (type)	Output Shape	Param #	Connected to Trainable
=== ===			
input_2 (InputLayer)	[(None, 136)]	0	[] Y
embedding_1 (Embedding)	(None, 136, 100)	9445000	['input_2 [0][0]'] Y
conv1d_7 (Conv1D)	(None, 136, 32)	12832	['embeddi ng_1[0][0]'] Y
conv1d_8 (Conv1D)	(None, 136, 32)	19232	['embeddi ng_1[0][0]'] Y
conv1d_9 (Conv1D)	(None, 136, 32)	25632	['embeddi ng_1[0][0]'] Y
concatenate_2 (Concatenate)	(None, 136, 96)	0	['conv1d_ 7[0][0]', Y 'conv1d_ 8[0][0]', 'conv1d_ 9[0][0]']
max_pooling1d_2 (MaxPooling1D)	(None, 68, 96)	0	['concate nate_2[0][0]'] Y
conv1d_10 (Conv1D)	(None, 68, 32)	12320	['max_poo ling1d_2[0][0]'] Y
conv1d_11 (Conv1D)	(None, 68, 32)	18464	['max_poo ling1d_2[0][0]'] Y
conv1d_12 (Conv1D)	(None, 68, 32)	24608	['max_poo ling1d_2[0][0]'] Y
concatenate_3 (Concatenate)	(None, 68, 96)	0	['conv1d_ 10[0][0]', Y 'conv1d_ 11[0][0]', 'conv1d_ 12[0][0]']
max_pooling1d_3 (MaxPooling1D)	(None, 34, 96)	0	['concate nate_3[0][0]'] Y
conv1d_13 (Conv1D)	(None, 34, 32)	12320	['max_poo ling1d_3[0][0]'] Y
flatten_1 (Flatten)	(None, 1088)	0	['conv1d_ 13[0][0]'] Y

```
 dropout_1 (Dropout)              (None, 1088)              0        ['flatten
_1[0][0]']              Y


 dense_2 (Dense)                  (None, 50)            54450        ['dropout
_1[0][0]']              Y


 dense_3 (Dense)                  (None, 20)             1020        ['dense_2
[0][0]']               Y


==============================================================================
================================
Total params: 9625878 (36.72 MB)
Trainable params: 9625878 (36.72 MB)
Non-trainable params: 0 (0.00 Byte)

_____
_____
```

In [68]:
```python
from sklearn.metrics import f1_score
from keras.callbacks import ModelCheckpoint
from keras.callbacks import EarlyStopping
from keras.callbacks import ReduceLROnPlateau

es = EarlyStopping(monitor='val_accuracy', mode='max', verbose=1,patience=
mc = ModelCheckpoint('/content/drive/MyDrive/Srisailam/best_model_1.h5', m
                     mode='max', save_best_only=True,verbose=1)
reduce_lr = ReduceLROnPlateau(monitor='val_accuracy', factor=0.8, patience

callback_list = [es,mc,reduce_lr]
```

In [69]:
```python
model.fit(train_padded,y_train,batch_size=64,epochs=25,verbose=1,validatic
          callbacks = callback_list)
```

```
Epoch 1/25
235/235 [==============================] - ETA: 0s - loss: 3.0017 - accuracy:
0.0573
Epoch 1: val_accuracy improved from -inf to 0.08520, saving model to /conten
t/drive/MyDrive/Srisailam/best_model_1.h5

/usr/local/lib/python3.10/dist-packages/keras/src/engine/training.py:3103: Us
erWarning: You are saving your model as an HDF5 file via `model.save()`. This
file format is considered legacy. We recommend using instead the native Keras
format, e.g. `model.save('my_model.keras')`.
  saving_api.save_model(
```

```
235/235 [==============================] - 24s 93ms/step - loss: 3.0017 - acc
uracy: 0.0573 - val_loss: 2.9627 - val_accuracy: 0.0852 - lr: 1.0000e-04
Epoch 2/25
235/235 [==============================] - ETA: 0s - loss: 2.8062 - accuracy:
0.1283
Epoch 2: val_accuracy improved from 0.08520 to 0.22560, saving model to /cont
ent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 9s 37ms/step - loss: 2.8062 - accu
racy: 0.1283 - val_loss: 2.4359 - val_accuracy: 0.2256 - lr: 1.0000e-04
Epoch 3/25
235/235 [==============================] - ETA: 0s - loss: 2.2481 - accuracy:
0.2492
Epoch 3: val_accuracy improved from 0.22560 to 0.34860, saving model to /cont
ent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 7s 28ms/step - loss: 2.2481 - accu
racy: 0.2492 - val_loss: 1.9057 - val_accuracy: 0.3486 - lr: 1.0000e-04
Epoch 4/25
235/235 [==============================] - ETA: 0s - loss: 1.8566 - accuracy:
0.3508
Epoch 4: val_accuracy improved from 0.34860 to 0.43920, saving model to /cont
ent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 5s 22ms/step - loss: 1.8566 - accu
racy: 0.3508 - val_loss: 1.6200 - val_accuracy: 0.4392 - lr: 1.0000e-04
Epoch 5/25
235/235 [==============================] - ETA: 0s - loss: 1.5009 - accuracy:
0.4667
Epoch 5: val_accuracy improved from 0.43920 to 0.57720, saving model to /cont
ent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 4s 18ms/step - loss: 1.5009 - accu
racy: 0.4667 - val_loss: 1.2483 - val_accuracy: 0.5772 - lr: 1.0000e-04
Epoch 6/25
235/235 [==============================] - ETA: 0s - loss: 1.2160 - accuracy:
0.5701
Epoch 6: val_accuracy improved from 0.57720 to 0.66300, saving model to /cont
ent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 5s 23ms/step - loss: 1.2160 - accu
racy: 0.5701 - val_loss: 1.0157 - val_accuracy: 0.6630 - lr: 1.0000e-04
Epoch 7/25
234/235 [=============================>.] - ETA: 0s - loss: 0.9944 - accuracy:
0.6489
Epoch 7: val_accuracy improved from 0.66300 to 0.71900, saving model to /cont
ent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 4s 16ms/step - loss: 0.9941 - accu
racy: 0.6490 - val_loss: 0.8333 - val_accuracy: 0.7190 - lr: 1.0000e-04
Epoch 8/25
235/235 [==============================] - ETA: 0s - loss: 0.8165 - accuracy:
0.7173
Epoch 8: val_accuracy improved from 0.71900 to 0.77560, saving model to /cont
ent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 4s 16ms/step - loss: 0.8165 - accu
racy: 0.7173 - val_loss: 0.6868 - val_accuracy: 0.7756 - lr: 1.0000e-04
Epoch 9/25
235/235 [==============================] - ETA: 0s - loss: 0.6794 - accuracy:
0.7658
Epoch 9: val_accuracy improved from 0.77560 to 0.81140, saving model to /cont
ent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 4s 17ms/step - loss: 0.6794 - accu
```

```
racy: 0.7658 - val_loss: 0.5934 - val_accuracy: 0.8114 - lr: 1.0000e-04
Epoch 10/25
235/235 [==============================] - ETA: 0s - loss: 0.5664 - accuracy:
0.8101
Epoch 10: val_accuracy improved from 0.81140 to 0.85300, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 4s 17ms/step - loss: 0.5664 - accu
racy: 0.8101 - val_loss: 0.4893 - val_accuracy: 0.8530 - lr: 1.0000e-04
Epoch 11/25
235/235 [==============================] - ETA: 0s - loss: 0.4773 - accuracy:
0.8434
Epoch 11: val_accuracy improved from 0.85300 to 0.86960, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 3s 15ms/step - loss: 0.4773 - accu
racy: 0.8434 - val_loss: 0.4227 - val_accuracy: 0.8696 - lr: 1.0000e-04
Epoch 12/25
235/235 [==============================] - ETA: 0s - loss: 0.4159 - accuracy:
0.8632
Epoch 12: val_accuracy improved from 0.86960 to 0.88500, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 8s 36ms/step - loss: 0.4159 - accu
racy: 0.8632 - val_loss: 0.3833 - val_accuracy: 0.8850 - lr: 1.0000e-04
Epoch 13/25
235/235 [==============================] - ETA: 0s - loss: 0.3538 - accuracy:
0.8853
Epoch 13: val_accuracy improved from 0.88500 to 0.89120, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 7s 30ms/step - loss: 0.3538 - accu
racy: 0.8853 - val_loss: 0.3425 - val_accuracy: 0.8912 - lr: 1.0000e-04
Epoch 14/25
235/235 [==============================] - ETA: 0s - loss: 0.3091 - accuracy:
0.9030
Epoch 14: val_accuracy improved from 0.89120 to 0.90520, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 4s 16ms/step - loss: 0.3091 - accu
racy: 0.9030 - val_loss: 0.3022 - val_accuracy: 0.9052 - lr: 1.0000e-04
Epoch 15/25
235/235 [==============================] - ETA: 0s - loss: 0.2765 - accuracy:
0.9112
Epoch 15: val_accuracy improved from 0.90520 to 0.91280, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 8s 34ms/step - loss: 0.2765 - accu
racy: 0.9112 - val_loss: 0.2861 - val_accuracy: 0.9128 - lr: 1.0000e-04
Epoch 16/25
235/235 [==============================] - ETA: 0s - loss: 0.2531 - accuracy:
0.9189
Epoch 16: val_accuracy did not improve from 0.91280
235/235 [==============================] - 6s 25ms/step - loss: 0.2531 - accu
racy: 0.9189 - val_loss: 0.2845 - val_accuracy: 0.9102 - lr: 1.0000e-04
Epoch 17/25
235/235 [==============================] - ETA: 0s - loss: 0.2239 - accuracy:
0.9289
Epoch 17: val_accuracy did not improve from 0.91280
235/235 [==============================] - 4s 16ms/step - loss: 0.2239 - accu
racy: 0.9289 - val_loss: 0.2778 - val_accuracy: 0.9040 - lr: 1.0000e-04
Epoch 18/25
235/235 [==============================] - ETA: 0s - loss: 0.2019 - accuracy:
```

```
0.9348
Epoch 18: val_accuracy improved from 0.91280 to 0.91840, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 5s 23ms/step - loss: 0.2019 - accu
racy: 0.9348 - val_loss: 0.2355 - val_accuracy: 0.9184 - lr: 8.0000e-05
Epoch 19/25
235/235 [==============================] - ETA: 0s - loss: 0.1820 - accuracy:
0.9421
Epoch 19: val_accuracy improved from 0.91840 to 0.92460, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 4s 16ms/step - loss: 0.1820 - accu
racy: 0.9421 - val_loss: 0.2294 - val_accuracy: 0.9246 - lr: 8.0000e-05
Epoch 20/25
235/235 [==============================] - ETA: 0s - loss: 0.1678 - accuracy:
0.9467
Epoch 20: val_accuracy did not improve from 0.92460
235/235 [==============================] - 3s 12ms/step - loss: 0.1678 - accu
racy: 0.9467 - val_loss: 0.2337 - val_accuracy: 0.9204 - lr: 8.0000e-05
Epoch 21/25
235/235 [==============================] - ETA: 0s - loss: 0.1598 - accuracy:
0.9489
Epoch 21: val_accuracy improved from 0.92460 to 0.92480, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 4s 17ms/step - loss: 0.1598 - accu
racy: 0.9489 - val_loss: 0.2090 - val_accuracy: 0.9248 - lr: 8.0000e-05
Epoch 22/25
235/235 [==============================] - ETA: 0s - loss: 0.1532 - accuracy:
0.9501
Epoch 22: val_accuracy improved from 0.92480 to 0.92660, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 4s 16ms/step - loss: 0.1532 - accu
racy: 0.9501 - val_loss: 0.2148 - val_accuracy: 0.9266 - lr: 8.0000e-05
Epoch 23/25
234/235 [=============================>.] - ETA: 0s - loss: 0.1424 - accuracy:
0.9554
Epoch 23: val_accuracy improved from 0.92660 to 0.92900, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 3s 13ms/step - loss: 0.1423 - accu
racy: 0.9555 - val_loss: 0.1998 - val_accuracy: 0.9290 - lr: 8.0000e-05
Epoch 24/25
235/235 [==============================] - ETA: 0s - loss: 0.1388 - accuracy:
0.9530
Epoch 24: val_accuracy improved from 0.92900 to 0.93160, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 3s 13ms/step - loss: 0.1388 - accu
racy: 0.9530 - val_loss: 0.1962 - val_accuracy: 0.9316 - lr: 8.0000e-05
Epoch 25/25
235/235 [==============================] - ETA: 0s - loss: 0.1300 - accuracy:
0.9571
Epoch 25: val_accuracy improved from 0.93160 to 0.93380, saving model to /con
tent/drive/MyDrive/Srisailam/best_model_1.h5
235/235 [==============================] - 3s 15ms/step - loss: 0.1300 - accu
racy: 0.9571 - val_loss: 0.2049 - val_accuracy: 0.9338 - lr: 8.0000e-05
```

Out[69]: <keras.src.callbacks.History at 0x78fd62b86320>

In [70]:
```python
def plot_history(history,accuracy_title,loss_title):
    # summarize history for accuracy
    plt.plot(model.history.history['accuracy'])
    plt.plot(model.history.history['val_accuracy'])
    plt.title(accuracy_title)
    plt.ylabel('Accuracy')
    plt.xlabel('Epoch')
    plt.legend(['Training', 'Validation'], loc='upper right')
    plt.show()

    # summarize history for loss
    plt.plot(model.history.history['loss'])
    plt.plot(model.history.history['val_loss'])
    plt.title(loss_title)
    plt.ylabel('Loss')
    plt.xlabel('Epoch')
    plt.legend(['Training', 'Validation'], loc='upper right')
    plt.show()
```

In [71]: `1` `plot_history(model,``'CNN Model Accuracy'``,``'CNN Model Loss'``)`

CNN Model Accuracy

CNN Model Loss
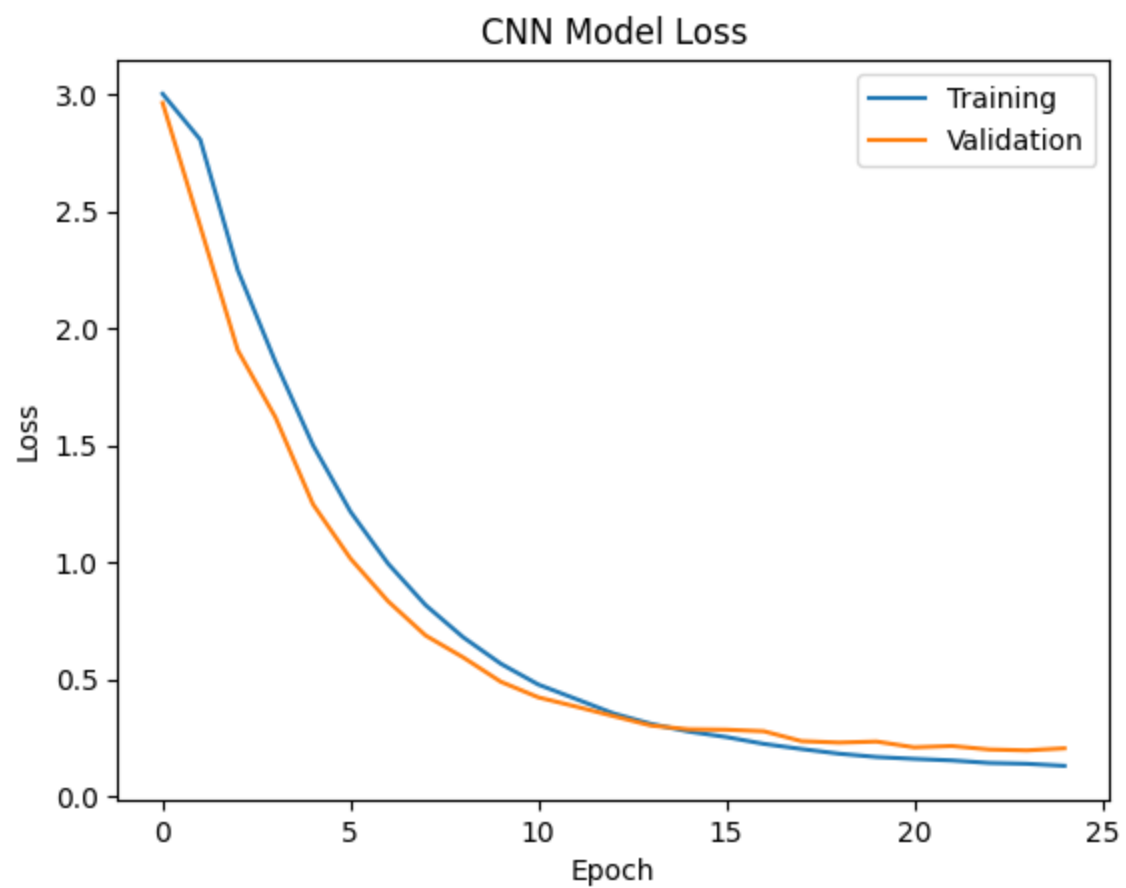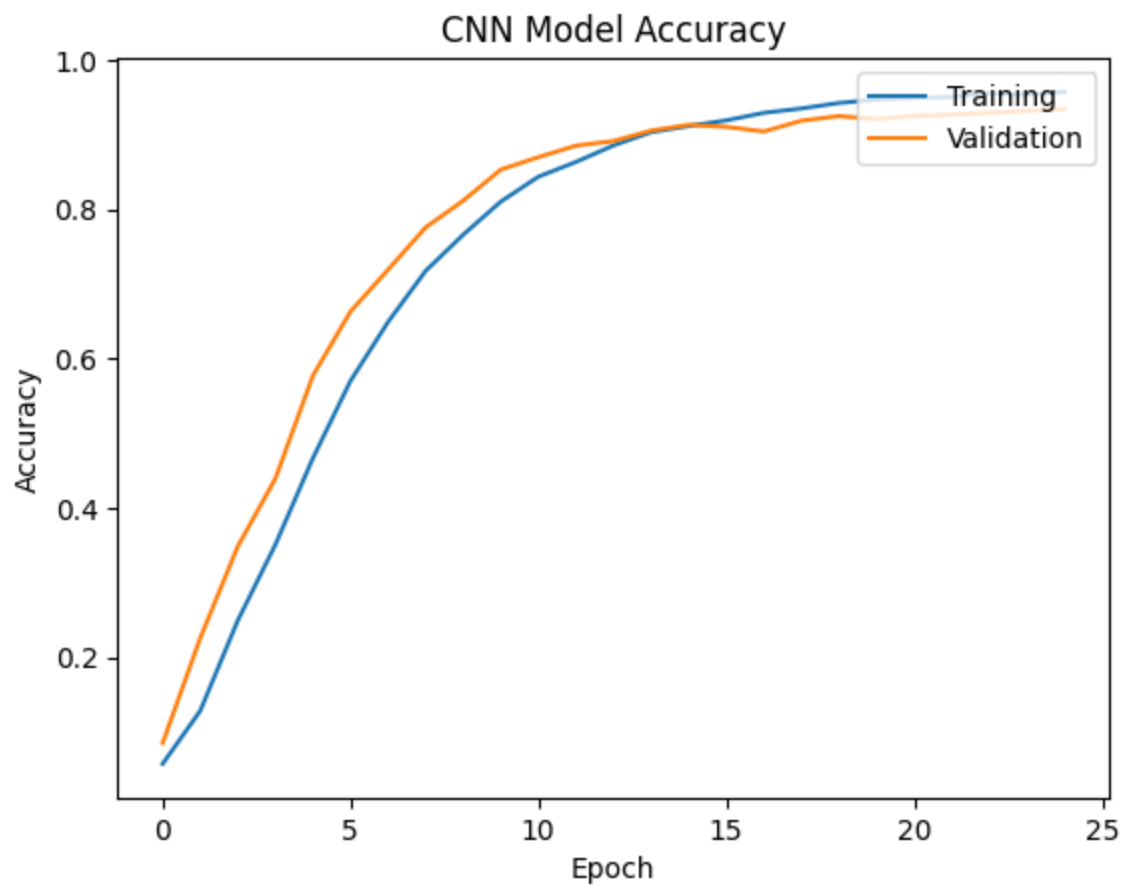
Machine Learning Models

In [72]:
```python
1  from sklearn.linear_model import LogisticRegression
2  from sklearn.ensemble import RandomForestClassifier
3  from sklearn.metrics import accuracy_score
```

In [73]:
```python
1  LR = LogisticRegression(random_state=0,max_iter=10000).fit(train_padded, 
```

In [74]:
```python
1  train_preds = LR.predict(train_padded)
2  train_accuracy = accuracy_score(np.argmax(y_train,axis=1), train_preds)
3  print('Train Accuracy Score using Logistic Regression is ',train_accuracy)
4
5  test_preds = LR.predict(test_padded)
6  test_accuracy = accuracy_score(np.argmax(y_test,axis=1), test_preds)
7  print('Test Accuracy Score using Logistic Regression is ',test_accuracy)
```

```
Train Accuracy Score using Logistic Regression is  0.16309928652397146
Test Accuracy Score using Logistic Regression is  0.1078
```

In [75]:
```python
1  RF = RandomForestClassifier(max_depth=4, random_state=0).fit(train_padded,
```

In [76]:
```python
1  train_preds = RF.predict(train_padded)
2  train_accuracy = accuracy_score(np.argmax(y_train,axis=1), train_preds)
3  print('Train Accuracy Score using Random Forest is ',train_accuracy)
4
5  test_preds = RF.predict(test_padded)
6  test_accuracy = accuracy_score(np.argmax(y_test,axis=1), test_preds)
7  print('Test Accuracy Score using Random Forest is ',test_accuracy)
```

```
Train Accuracy Score using Random Forest is  0.15689804627592185
Test Accuracy Score using Random Forest is  0.1286
```

In [80]:
```python
1  #Summarizing the results obtained from each trained model
2
3  from prettytable import PrettyTable
4  x = PrettyTable()
5  x.field_names = ["Model", "Train Accuracy", "Test Accuracy"]
6  x.add_row(["CNN",0.957,0.933])
7  x.add_row(["Logistic Regression",0.163,0.10])
8  x.add_row(["Random Forest",0.156,0.128])
9  print(x)
```

```
+---------------------+----------------+---------------+
|        Model        | Train Accuracy | Test Accuracy |
+---------------------+----------------+---------------+
|         CNN         |     0.957      |     0.933     |
| Logistic Regression |     0.163      |      0.1      |
|    Random Forest    |     0.156      |     0.128     |
+---------------------+----------------+---------------+
```

Observations:

We can observe that the CNN has outperformed both logistic regression and Random Forest in terms of accuracies.

In [78]:
```python
from keras.models import load_model
best_model = load_model('/content/drive/MyDrive/Srisailam/best_model_1.h5'
```

In [79]:

```python
#Predictions on random 20 points in test data
import random
for _ in range(20):
  i = random.randint(0,X_test.shape[0])
  print(X_test[i])
  print('\n')

  #Predicting the labels using the best model obtained above.
  pred_label = np.argmax(best_model.predict(test_padded[i].reshape(1,-1)))
  actual_label = np.argmax(y_test[i])
  print('Actual Label - ',le.classes_[actual_label])
  print('Predicted Label - ',le.classes_[pred_label])
  print(100*'*')
```

descartes etl army mil tecsun1 tec army mil wmich edu tecsun1 tec army mil on
ion cmu edu tecsun1 tec army mil onion cmu edu cmu edu tecsun1 tec army mil d
escartes etl army milwho s next mormons and jewscantaloupe srv cmu edu sci ta
lk politics talk religion misc cantaloupe srv cmu edu rochester udel gatech h
owland reston ans net usc utexas edu uunet iris mbvlab wpafb mil blackbird af
it mil afterlife tecsun descartes riggs sci skeptic talk politics misc talk r
eligion misc misc legal message apr gmt followup sci skeptic talk politics mi
sc talk religion misc misc legal lnk corporation riverdale nntp posting desca
rtes tec army mil was slightly surprised see the guns roston open here but ha
s his posting record and have mine although usually more polite than not afra
id controversy either this looks like argument that started just contrary pan
der the sci skeptic masses will certainly bow out schedule for the next two w
eeks does not read april through debate roston the true meaning the declarati
on independence and the bill rights include long discourse the validity claim
s the declaration independence have crossposted this misc legal that real law
yers can pass judgement legal arguments article article one thing that should
made clear that neither the fbi nor the batf responsible for what happened ye
sterday one can argue about the initial raid but would worth mentioning befor
e the facts get lost that the branch davidians were tipped off that the batf
was coming during the initial raid the branch davidians opened fire first was
waiting some some not intelligent person bring this look this country laws qu
ote piece parchment that many seem think little amendment the right the peopl
e secure their persons houses papers and effects against unreasonable searche
s and seizures shall not violated and warrants shall issue but upon probable
cause supported oath affirmation and particularly describing the place search
ed and the persons things seized the first question did the notip off call in
dicate the branch davidians that knock warrant was use and they could therefo
re presumably into weapons free once the compound property was entered the se
cond question what makes think that the davidians actions would have been dif
ferent had another type warrant been use would like see both these questions
answered before seeking judgment any claims about knock warrant clear violati
on the amendment okay what about the fact that they were tipped off they shou
ld not have opened fire right wrong think about drug dealer and competition w
ants away with they call and tell that the feds are their way with knock warr
ant being moronic sheep wait with our guns holstered now instead the feds com
es competition and are history the only acceptable answer knock warrant blazi
ng guns may sound paranoid but our government out control and killing few fed
eral officers make knock some sense back into spurious analogy especially for
one skeptical the davidians were not drug dealers and all accounts feared one
but the feds any evidence the contrary when dangerous faction starts opinions
not represent with wrath and vengeance their hearts those employer solemn lea
gue and cov nant bound any government agency ruin slaughter and confound turn
religion fable and turn the government babel the law disgrace the gown corrup
t the senate rob the crown sacrifice old england glory and make her infamous
story when such tempest shook the land how could unguarded virtue stand

1/1 [==============================] - 0s 163ms/step
Actual Label - talk.politics.misc
Predicted Label - talk.religion.misc
****
****

unisql uucp unisql uucp unisql uucp srgenprp srgenprpwhere can i get a new yo
rk taxicantaloupe srv cmu edu magnesium club cmu edu news sei cmu edu cis ohi
o state edu pacific mps ohio state edu zaphod mps ohio state edu utexas edu u
nisql wrat rec autos message apr gmt unisql inc austin texas usa article dork

```
1/1 [==============================] - 0s 18ms/step
Actual Label -  rec.autos
Predicted Label -  rec.autos
**********************************************************************
**********************
```

sciences sdsu edu larc sdsu edu eagle lerc nasa gov eagle lerc nasa gov bach
lerc nasa gov sciences sdsu edu ucssun1 sdsu eduelectric power line balls sci
electronics message apr gmt article larc babb followup sci electronics usa sd
su larc nntp posting larc sdsu edu article got question from dad which really
can not answer and would appreciate some net wisdom his question about some d
iameter balls which are attached electric power lines his area seen half doze
n between two poles neither have any experience with electric power distribut
ion only guess was that they may capacitive device equalize the inductance th
e grid but why many between two poles will bet there runway nearby sdsu larc

```
1/1 [==============================] - 0s 20ms/step
Actual Label -  sci.electronics
Predicted Label -  sci.electronics
**********************************************************************
**********************
```

umbc edu umbc7 umbc edu umbc edulooking for an r5 xserver for hp9000 385canta
loupe srv cmu edu rochester udel news udel edu darwin sura net haven umd edu
news umbc edu nobody comp windows apr university maryland baltimore county ca
mpus message nntp posting umbc umbc edu auth banz anyone know where can find
one binaries are nice but source would thanks advance give people alternative
microsoft will have been greater good yes are your father ibm presentation fo
se

```
1/1 [==============================] - 0s 18ms/step
Actual Label -  comp.windows.x
Predicted Label -  comp.windows.x
**********************************************************************
**********************
```

umd edu mimsy umd edu cmu edu network cyberia win net mimsy umd edu cyberia w
in net umd educlinton wants national id card aka ussr style internal passport
cantaloupe srv cmu edu talk politics talk politics alt fan rush talk politics
guns talk politics drugs alt fan rush limbaugh message apr gmt followup talk
politics guns the reality liberation front how bout embed the card the forhea
d everyones skull not lose without being already dead close birth implant sma
rt chip just behind the ear under the skin but above the skull incase hypo al
lergenic high carbon content glass this chip would reprogrammed age with the
pertinent medical correctional taxational data behave yourself will input int
o your permenant record you forgot the part about encasing small shaped charg
e that anyone tries tamper with explodes and kills you and the shaped charge
can set off remote control but only you get out line properly patriotic citiz
ens have nothing fear born die your laws off brain what funny bout peace love
and understanding this sig contains animal products and was not tested animal
s congress shall make law respecting establishment religion prohibiting the f
ree exercise thereof abridging the freedom speech the press the right the peo
ple peaceably assemble and petition the government for redress grievances the
constitution the united states america

```
1/1 [==============================] - 0s 20ms/step
```

```
Actual Label -  talk.politics.guns
Predicted Label -  talk.politics.guns
************************************************************************
**********************
```

aurora alaska edu aurora alaska edu news duc auburn edu raven alaska edu news duc auburn edu eng auburn edu eng auburn edu acad3 alaska edusolar sail datac antaloupe srv cmu edu crabapple srv cmu edu andrew cmu edu news sei cmu edu c is ohio state edu zaphod mps ohio state edu uwm edu ogicse news washington ed u raven alaska edu aurora alaska edu nsmca sci space message apr gmt article aurora apr university alaska fairbanks nntp posting acad alaska edu article l ooking for any information concerning projects involving solar sails understa nd that the jpl did extensive study the subject back the late but having trou ble gathering such information are there any groups out there currently invol ved such project know someone had long talks about solar sails early this yea r and late last year also about solar sailing not sure who captured possible think was one the regulars who had most all the data think started the latest round the late last year round but the topic has been around here off and for year two not high just jacked

```
1/1 [==============================] - 0s 20ms/step
Actual Label -  sci.space
Predicted Label -  sci.space
************************************************************************
**********************
```

ccc amdahl juts ccc amdahl byuvm bitnet juts ccc amdahl byuvm bitnet byuvm bi tnet byuvm bitnetrw is robert weiss the only orthodox christiancantaloupe srv cmu edu magnesium club cmu edu news sei cmu edu ece cmu edu europa eng gtefsd com howland reston ans net usc utexas edu uunet amdahl juts agr talk religion misc message apr gmt reply amdahl corporation sunnyvale article you keep maki ng references orthodox belief and saying things like held that what exact bod y theology are you drawing for what you call orthodox who that holds that mea nt what you said meant whenever your personal interpretation biblical passage s challenged your only response seems that one needs merely look the bible or der see the truth but what those who see biblical things differently from you are simply assume that you are the only one who really understands just curio us when refers the orthodox talking about the historical position the christi an faith such things are derived from biblcal texts through the centuries the apocolic fathers the faith you are right that people read things differently the bible and this alright parts like parables and such forth however when co mes the essential doctrines the historical orthodox christan beliefs there on ly one correct way read for example either the doctrine the trinity true fals e yes people read the texts differently but only one position true they both cannot according the text the doctrine true and has always existed therefore when people like come along with vision and thinks can undo centuries doctrin e that supported the bible people consider him cult

```
1/1 [==============================] - 0s 21ms/step
Actual Label -  talk.religion.misc
Predicted Label -  talk.religion.misc
************************************************************************
**********************
```

uiuc edu geneva rutgers edu geneva rutgers edu aramis rutgers edu uiuc educel l church discussion groupcantaloupe srv cmu edu crabapple srv cmu edu andrew cmu edu news sei cmu edu ece cmu edu europa eng gtefsd com howland reston ans net usc rutgers igor rutgers edu geneva rutgers edu christian soc religion ch

ristian message apr gmt university illinois dept comp sci beginning mail disc
ussion group about cell churches you are follower and are cell church church
that transitioning cell church just interested learning more about cell churc
hes send mail fool who gives what cannot keep gain what cannot lose

1/1 [==============================] - 0s 19ms/step
Actual Label - soc.religion.christian
Predicted Label - soc.religion.christian
****
****
scicom alphacdc com scicom alphacdc com cmu edu cmu edu flb optiplan scicom a
lphacdc comfirst spacewalkcantaloupe srv cmu edu rochester udel louie newsser
ver jvnc net howland reston ans net usenet ins cwru edu magnus acs ohio state
edu csn csn scicom wats sci space message apr gmt sci alpha science computer
network denver article one time there was speculation that the first spacewal
k was staged fake has any evidence support contradict this claim emerged was
this claim perhaps another fevered cold war hallucination for one would avid
reader sci space ussr what really happened item extract

1/1 [==============================] - 0s 18ms/step
Actual Label - sci.space
Predicted Label - sci.space
****
****
sdc boeing fury boeing com news washington edu mojo intellection sdc boeing m
ojo intellection intellection news washington edu stein washington edu sun co
m sdc boeing sdc boeingproblem with r4 releasecantaloupe srv cmu edu comp win
dows comp windows comp windows comp windows intrinsics message may gmt follow
up comp windows boeing computer services seattle nntp posting zulu article ar
ticle unidentified external symbol use lxaw lxext lxt not this order but prop
er order wondering there has some changes link libraries should include some
more libraries any help this matter will very useful and highly appreciated g
et this too you put bstatic into your options goes away things seem run witho
ut must not called very often from the sunos sun fixed shared library bug whi
ch conflicts with the way builds the shared library causing these symbols not
ably undefined when building some compiling bstatic bdynamic appears work sol
ve the problem you are using openwindows please contact your local sun office
and request the following patch description openwindows libxt jumbo patch ope
nwindows undefined symbols when using shared lib source patch for use with th
e mit libraries was developed retrofits into some fixes made get around this
problem the patch export contrib sunos patch_version client server tech servi
ces boeing computer services box seattle

1/1 [==============================] - 0s 21ms/step
Actual Label - comp.windows.x
Predicted Label - comp.windows.x
****
****
zoo toronto edu zoo toronto edu aio jsc nasa gov tm0006 lerc nasa gov cbl umd
edu cbl umd edu starburst umd edu zoo toronto eduspace station redesign jsc a
lternative 4sci space cantaloupe srv cmu edu magnesium club cmu edu news sei
cmu edu cis ohio state edu zaphod mps ohio state edu malgudi oar net news ans
net europa eng gtefsd com howland reston ans net usc utexas edu utnut utzoo h
enry message sat apr gmt toronto zoology article the only thing that scares t

he part about simply strapping ssme and nosecone and just launching have this vision something going terribly wrong with the launch resulting the complete loss the new modular space station does not make whole lot difference actually since they were not building spares the station hardware anyway least this only one launch fail svr resembles high speed collision toronto zoology between svr and sunos utzoo henry

```
1/1 [==============================] - 0s 18ms/step
Actual Label -  sci.space
Predicted Label -  sci.space
**************************************************************************************************
```

fsphy1 physics fsu edu world std mailer fsu edu fsphy1 physics fsu edu fsphy1 physics fsu edu world std world std fsphy1 physics fsu edumiscelaneous soon to have baby questionssci med cantaloupe srv cmu edu crabapple srv cmu edu ece cmu edu europa eng gtefsd com howland reston ans net wupost sdd com saimiri primate wisc edu eng ufl edu usenet ufl edu mailer fsu edu fsphy physics fsu edu picl florida state university school higher thought wed apr gmt news vax vms vnews message reply all article writes are about have our first baby and have few questions that dont seem able get answered our satisfaction reguarding having baby boy circumsized what are the medical pros and cons all have heard its the parents unfortonately that truly about the best summation the research that there advantages stated circumcison included probably prevention penile cancer simplicity personal hygiene prevention urinary tract infections and prevention unretractible foreskin disadvantages include infection from the procedure pain etc apologize trying pull this off the top head will post what discovered research did paper the topic research class nursing school really decision that the parents some parents use the reasoning that they will look like daddy and like their friends justification there nothing wrong with this just sure what you want since rather difficult uncircumcise male although major surgical procedure exists how about the pregnant woman sitting tub water have heard stories infection etc how about after the water has broken long your membranes have not broken and you have not had any problems with your pregnancy should sit tub water however would recommend using your own bathtub your own home nearly impossible guarantee the cleanliness and safety public hot tubs nice warm bath can very relaxing especially your back killing you and would possibly advisable avoid bubble bath soap esp you are prone yeast infection hope these tips help you some

```
1/1 [==============================] - 0s 18ms/step
Actual Label -  sci.med
Predicted Label -  sci.med
**************************************************************************************************
```

venus iucf indiana edu usenet ucs indiana edu usenet ucs indiana edu venus iucf indiana edu vax5 cit cornell edu laura harvard edu stafford winona msus edu stafford winona msus edu vax2 winona msus edu vax5 cit cornell edushaft drives and wheeliesrec motorcycles cantaloupe srv cmu edu magnesium club cmu edu news sei cmu edu cis ohio state edu pacific mps ohio state edu zaphod mps ohio state edu sol ctr columbia edu usenet ucs indiana edu venus iucf indiana edu brown message followup rec motorcycles news vax vms vnews nntp posting venus iucf indiana edu reply iucf rec tue apr gmt article writes apr gmt possible wheelie motorcycle with shaft drive yes but the rear wheel comes off the ground not the front see just hops into the air figure you can wheelies with shaft drive bike had bmw that was wheelie monster course did not have the initial

power burst just twist into the air had pop the clutch also had replace front
fork seals few times well the fairing bit heavy slamming down onto those litt
le stantion tubes all the time but let give you fair trashed the ring pinion
gear the final drive doing wheelies and this was cheap fix either there some
kind slip device the shaft prevent from breaking unfortunately did not save t
he gears the topic wheelies the other day saw kid big hurricane stoppy rear w
heelie man had the rear end this bike about feet off the ground traffic light
not recommend these activities anymore but looked damn impressive you can not
keep both tires the ground least have pointed that direction cheers


```
1/1 [==============================] - 0s 18ms/step
Actual Label -  rec.motorcycles
Predicted Label -  rec.motorcycles
********************************************************************************
**********************
```

porto inescn athena mit edu porto inescn export lcs mit edu porto inescnsubsc
ribecantaloupe srv cmu edu magnesium club cmu edu news sei cmu edu ece cmu ed
u europa eng gtefsd com howland reston ans net usc enterpoop mit edu internet
news mail gateway comp windows apr the internet message nntp posting enterpoo
p mit edu please subscribe


```
1/1 [==============================] - 0s 19ms/step
Actual Label -  comp.windows.x
Predicted Label -  comp.windows.x
********************************************************************************
**********************
```

ing puc tolten puc tolten puc bcsaic boeing bcsaic boeing ing pucfile sharing
magneto opticalscantaloupe srv cmu edu comp sys mac comp sys mac comp sys mac
system comp sys mac hardware cantaloupe srv cmu edu rochester udel darwin sur
a net haven umd edu uunet puc pablo message followup comp sys mac system comp
sys mac hardware malloco ing puc tin mon apr gmt jann vanover worker has just
attached magneto optical drive his mac for him however tried turn file sharin
g but would not work had some message about not all volumes are shareable has
anyone had success sharing mos please tell how needs insert the before filesh
aring get turned and the only one that can see the disk the owner the machine
bad you can not eject the disk until turned off hope that helps pablo millan
mis opiniones son mias email santiago chile research what doing when not know
what doing


```
1/1 [==============================] - 0s 18ms/step
Actual Label -  comp.sys.mac.hardware
Predicted Label -  comp.sys.mac.hardware
********************************************************************************
**********************
```

optilink com optilink com netcom panix ads ads ads com beowulf jpl nasa gov a
dsa loathesome subjectcantaloupe srv cmu edu crabapple srv cmu edu ece cmu ed
u europa eng gtefsd com emory swrinde utexas edu uunet optilink cramer politi
cs talk politics misc message apr gmt optilink corporation petaluma article j
ust curious how many times have you had sex with boys why you care total stra
nger asked you how often you had sex would you answer booz allen hamilton inc
this typical homosexuals constantly making excuses for child molesters uunet
pyramid optilink cramer opinions all mine relations between people mutual con
sent not all

```
1/1 [==============================] - 0s 19ms/step
Actual Label -  talk.politics.misc
Predicted Label -  talk.politics.misc
*****************************************************************************
***********************
acsu buffalo edu acsu buffalo edu hydra gatech edu acsu buffalo edu hydra gat
ech edu prism gatech edu acsu buffalo edu ubms buffalo edu ubvms buffalo edug
oalie mask updatecantaloupe srv cmu edu das news harvard edu noc near net uun
et zaphod mps ohio state edu acsu buffalo edu hammerl rec sport hockey messag
e apr gmt nntp posting lictor acsu buffalo edu article here are the results a
fter three days voting remember pts for for and for also you can still turn v
otes and the guy not regular goalie retired please include the team thanks fo
r your time and keep sending those votes soderstrom plays with philly but doe
s not have moulded mask got the helmet and cage variety white least that what
wore thirteen hours ago some days have remind him not claude lemieux top scor
er for the devils but known for taking dumb penalties
```

```
1/1 [==============================] - 0s 21ms/step
Actual Label -  rec.sport.hockey
Predicted Label -  rec.sport.hockey
*****************************************************************************
***********************
cbnewsj att cbnewsj att news ysu edu worwig atlantaga ncr com worwig atlantag
a ncr com atlantaga ncr com torreys attshaft drives and wheeliesrec motorcycl
es cantaloupe srv cmu edu crabapple srv cmu edu andrew cmu edu news sei cmu e
du cis ohio state edu zaphod mps ohio state edu caen msuinfo uchinews att out
cbnewsj car apr gmt message another one bites the dust article possible wheel
ie motorcycle with shaft drive imposible due the shaft effect the centripital
effects the rotating shaft counteract any tendency for the front wheel lift o
ff the ground well last two motorcycles have been shaft driven and they will
wheelie the rear gear does climb the ring gear and lift the rear which gives
odd feel but still wheelies feeding frenzy notis april rec moto and the newbi
es are bitin catch all you want will make more
```

```
1/1 [==============================] - 0s 21ms/step
Actual Label -  rec.motorcycles
Predicted Label -  rec.motorcycles
*****************************************************************************
***********************
syl nec research nec research nec acs ucalgary myrddin imat myrddin imat seah
unt imat acs ucalgary acs ucalgary syl necmaxima chain waxrec motorcycles nec
systems laboratory inc fri apr gmt article article chain lubed wheel clean af
ter careful was similarly impressed with first but recently have started find
ing rust spots the chain went back least the sucker never rusted when was usi
ng the bluegoo spray the chain wax onto the rollers and sideplates occassiona
lly and rust will not problem later dod fxwg wide glide red lady baby bike no
w why would nec needs name agree with any this anyway was raised pack wild co
rn dogs
```

```
1/1 [==============================] - 0s 20ms/step
Actual Label -  rec.motorcycles
Predicted Label -  rec.motorcycles
*****************************************************************************
```

```
**********************
```

space alcbel alcbel news arc nasa gov alcbel space alcbel space alcbelstateme
nt by nasa administrator daniel s g news harvard edu noc near net howland res
ton ans net zaphod mps ohio state edu utexas edu uunet pipex uknet mcsun alcb
el btmpj devdjn sci space message apr gmt reply alcatel bell telephone nntp p
osting this man nasa administrator then god save nasa course the record unriv
aled there only one shuttle furthermore there only likely one shuttle now tha
t hermes and are effectively cancelled these officials should spend more thei
r time explaining their european and asian partners how are expected believe
them when their paymasters change their minds major international projects ev
erytime new administration takes office considering the major impacts this ha
s the european and asian industry also appreciated how this affects american
industry course talking about space stattion freedom alcatel bell telephone b
erkenrodelei hoboken belgium

1/1 [==============================] - 0s 18ms/step
Actual Label - sci.space
Predicted Label - sci.space
****
****
```