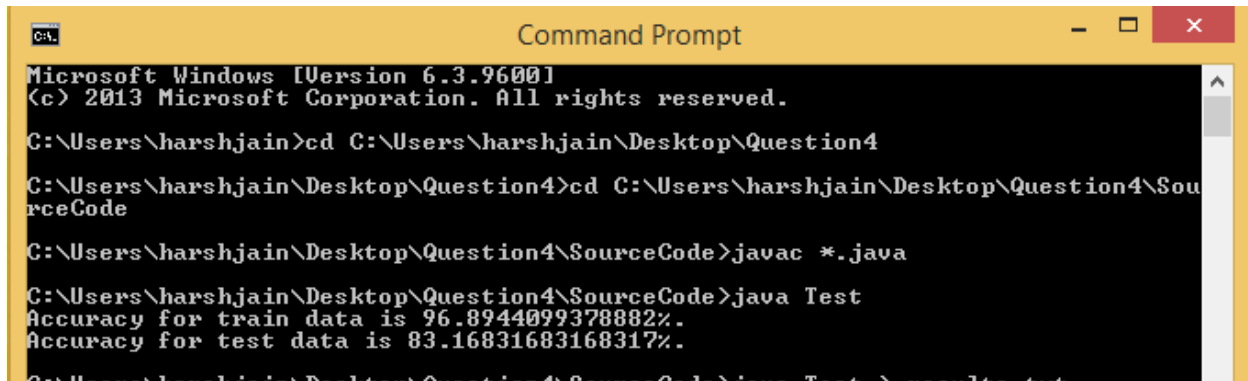


Following are the results for training data and test data:



```
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\Users\harshjain>cd C:\Users\harshjain\Desktop\Question4

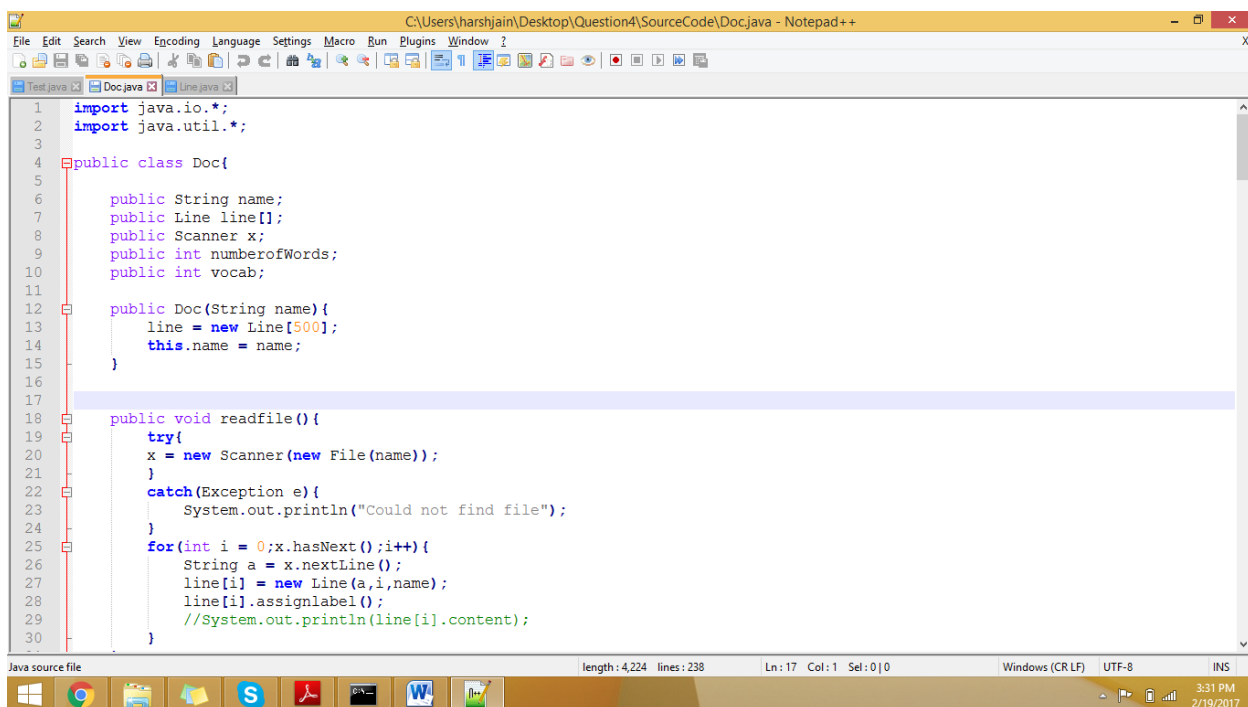
C:\Users\harshjain\Desktop\Question4>cd C:\Users\harshjain\Desktop\Question4\SourceCode

C:\Users\harshjain\Desktop\Question4\SourceCode>javac *.java

C:\Users\harshjain\Desktop\Question4\SourceCode>java Test
Accuracy for train data is 96.8944099378882%.
Accuracy for test data is 83.16831683168317%.
```

Three classes created to implement the text classification : Doc.class, Line.class and Test.class. SourceCode attached in zip file. Following is the code screenshot:

Doc.class



```
import java.io.*;
import java.util.*;

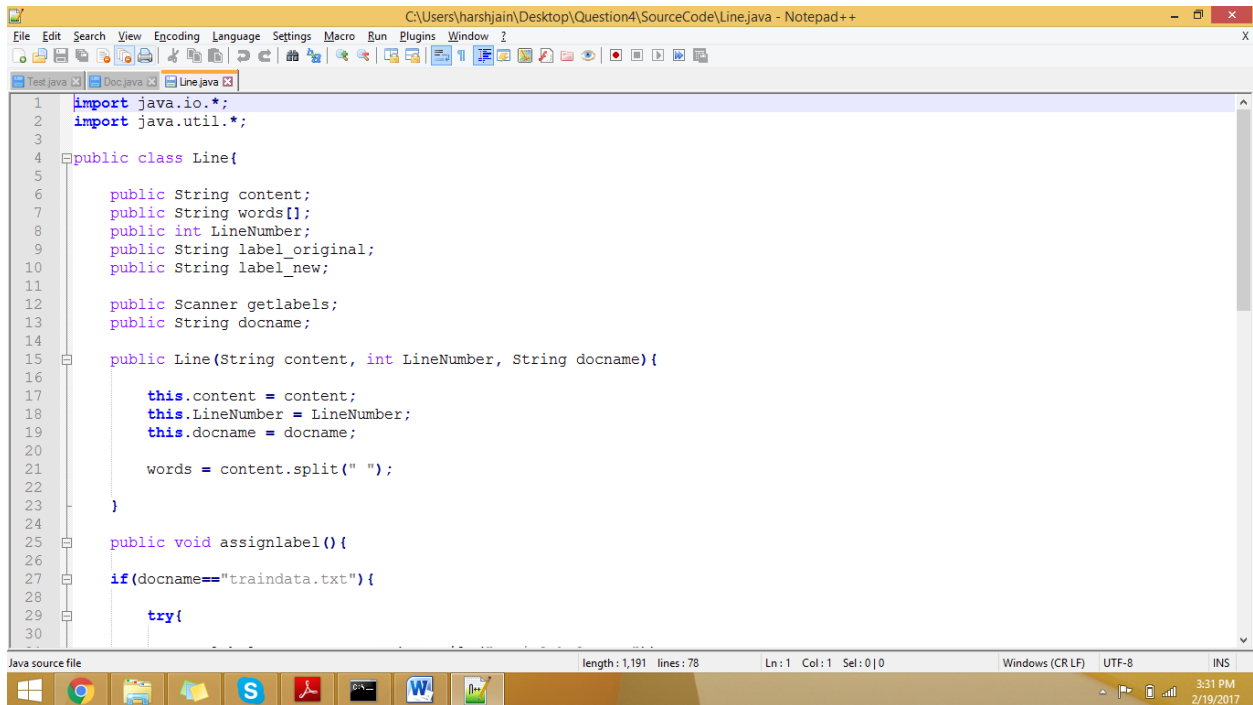
public class Doc{

    public String name;
    public Line line[];
    public Scanner x;
    public int numberOfWords;
    public int vocab;

    public Doc(String name){
        line = new Line[500];
        this.name = name;
    }

    public void readFile(){
        try{
            x = new Scanner(new File(name));
        }
        catch(Exception e){
            System.out.println("Could not find file");
        }
        for(int i = 0; x.hasNext(); i++){
            String a = x.nextLine();
            line[i] = new Line(a, i, name);
            line[i].assignlabel();
            //System.out.println(line[i].content);
        }
    }
}
```

Line.class

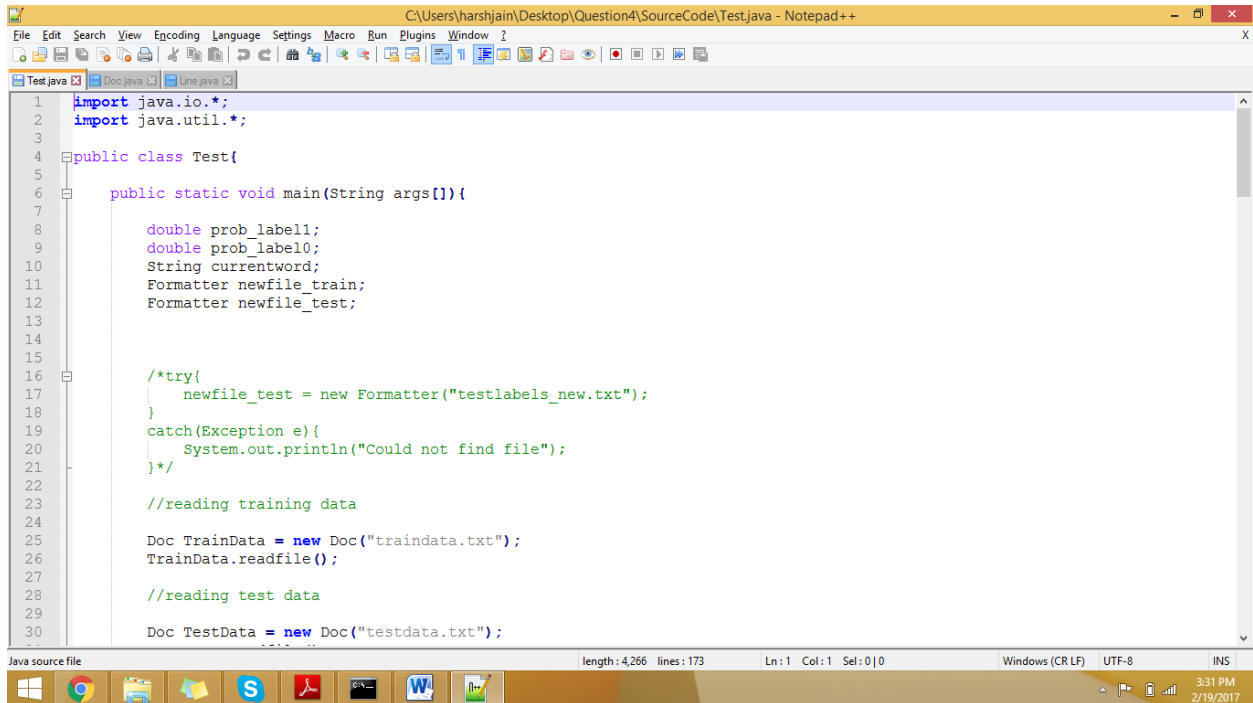


The screenshot shows a Notepad++ window titled "C:\Users\harshjain\Desktop\Question4\SourceCode\Line.java - Notepad++". The editor contains the following Java code for the Line class:

```
1  import java.io.*;
2  import java.util.*;
3
4  public class Line{
5
6      public String content;
7      public String words[];
8      public int LineNumber;
9      public String label_original;
10     public String label_new;
11
12     public Scanner getlabels;
13     public String docname;
14
15     public Line(String content, int LineNumber, String docname){
16
17         this.content = content;
18         this.LineNumber = LineNumber;
19         this.docname = docname;
20
21         words = content.split(" ");
22     }
23
24     public void assignlabel(){
25
26         if(docname=="traindata.txt"){
27
28             try{
29
30
```

The status bar at the bottom indicates "Java source file", "length: 1,191 lines: 78", "Ln: 1 Col: 1 Sel: 0 | 0", "Windows (CR LF)", "UTF-8", "INS", and the system clock shows "3:31 PM 2/19/2017".

Test.class



The screenshot shows a Notepad++ window titled "C:\Users\harshjain\Desktop\Question4\SourceCode\Test.java - Notepad++". The editor contains the following Java code for the Test class:

```
1  import java.io.*;
2  import java.util.*;
3
4  public class Test{
5
6      public static void main(String args[]){
7
8          double prob_label1;
9          double prob_label0;
10         String currentword;
11         Formatter newfile_train;
12         Formatter newfile_test;
13
14
15
16         /*try{
17             newfile_test = new Formatter("testlabels_new.txt");
18         }
19         catch(Exception e){
20             System.out.println("Could not find file");
21         }*/
22
23         //reading training data
24
25         Doc TrainData = new Doc("traindata.txt");
26         TrainData.readfile();
27
28         //reading test data
29
30         Doc TestData = new Doc("testdata.txt");
```

The status bar at the bottom indicates "Java source file", "length: 4,266 lines: 173", "Ln: 1 Col: 1 Sel: 0 | 0", "Windows (CR LF)", "UTF-8", "INS", and the system clock shows "3:31 PM 2/19/2017".