

# README:

Yijie Feng

JDK Version: 1.7.0

## Compile and Run Naive Bayes Text Classification Program:

1. run javac to compile

```
javac *.java
```

2. run program

```
java Main [input file] [N] [stopwords file]
```

- input file: the text corpus
- N: the number of entries to use for the training set (first N entries are taken)
- stopwords file: text file containing a list of stopwords (separated by spaces)

*example:*

```
java Main ./tinyCorpus.txt 5 ./stopwords.txt
```

---

## NOTES:

- stopwords are read in as input from a file called stopwords.txt in the root of the project; stopwords can be changed by replacing this file with a new stopwords.txt file
- I added some additional words to the stopwords file so that the program would match the trace shown on the course website