**Language used: Java**

Unzip and Open the Command Prompt (Terminal at the PartII folder).

The Multinomial Naïve Bayes algorithm was implemented using Java in Eclipse IDE.

The program will compile in the IDE with no errors. If IDE is not used, can compile using the

Following command in the cmd/terminal:

Source folder contains all required java files

1) Javac \*.java //will compile all the .java files

**How to run:**

A folder can be created and any 5 classes (i.e., sub folders) can be selected from the training dataset and copied into this folder. This will be the training set and this folder’s path is to be given as first command line argument.

Similarly another folder created as the test folder should contain the corresponding 5 classes’ test data and this folder’s path is given as second command line argument.

The code is then run. The main () is inside Bayes.java.

The project will output the precision, recall and F-score of every class and total accuracy.

Run the following command:

**java Bayes K:\\Dhwani\\1stSem\\ML\\Ass4\\Part2\\20news-bydate\\20news-bydate-train**

**K:\\Dhwani\\1stSem\\ML\\Ass4\\Part2\\20news-bydate\\20news-bydate-test**

**Sample Output:**

**1)**

Classes:

alt.atheism

comp.graphics

rec.autos

sci.crypt

talk.politics.guns

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Confusion Metrics\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alt.atheism comp.graphics rec.autos sci.crypt talk.politics.guns Total

alt.atheism 301 4 6 2 14 | 327

comp.graphics 4 372 8 20 4 | 408

rec.autos 14 12 380 8 16 | 430

sci.crypt 4 18 6 371 10 | 409

talk.politics.guns 14 0 12 20 342 | 388

-----------------------------------------------------------------

Total 337 406 412 421 386

-----------------------------------------------------------------

Class : alt.atheism

Precision : 0.8931750741839762

Recall : 0.9204892966360856

F-score : 0.9066265060240963

Class : comp.graphics

Precision : 0.916256157635468

Recall : 0.9117647058823529

F-score : 0.9140049140049141

Class : rec.autos

Precision : 0.9223300970873787

Recall : 0.8837209302325582

F-score : 0.9026128266033253

Class : sci.crypt

Precision : 0.8812351543942993

Recall : 0.9070904645476773

F-score : 0.8939759036144577

Class : talk.politics.guns

Precision : 0.8860103626943006

Recall : 0.8814432989690721

F-score : 0.883720930232558

Total Accuracy:94.74248927038627

**2)**

Classes:

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

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Confusion Metrics\*\*\*\*\*\*\*\*\*\*\*\*\*\*

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 Total

alt.atheism 247 2 4 0 0 0 0 2 2 0 2 0 2 18 6 14 4 18 10 94 | 425

comp.graphics 0 307 132 22 26 76 10 4 2 0 0 4 44 14 12 6 2 2 4 6 | 673

comp.os.ms-windows.misc 0 0 14 2 2 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | 22

comp.sys.ibm.pc.hardware 0 28 288 305 74 26 86 4 0 0 0 2 58 4 0 0 0 0 0 0 | 875

comp.sys.mac.hardware 0 16 36 54 283 4 46 0 0 2 0 4 18 0 2 0 0 0 2 0 | 467

comp.windows.x 2 48 188 12 8 321 10 4 0 6 0 6 6 2 6 2 0 6 2 0 | 629

misc.forsale 2 0 4 12 4 0 257 8 4 2 0 4 6 0 0 0 2 2 0 0 | 307

rec.autos 0 2 4 4 10 0 44 352 36 4 0 2 18 10 0 0 2 0 0 0 | 488

rec.motorcycles 0 4 2 0 2 4 8 10 363 0 0 0 6 0 0 0 2 2 0 2 | 405

rec.sport.baseball 0 0 6 2 2 2 2 4 4 360 6 0 0 6 0 0 0 2 0 0 | 396

rec.sport.hockey 4 0 10 6 12 2 16 6 2 40 387 2 6 8 4 0 2 2 0 0 | 509

sci.crypt 2 28 36 8 14 18 2 2 4 2 2 372 76 0 8 2 6 8 8 0 | 598

sci.electronics 2 6 4 48 24 0 16 12 2 0 0 4 259 8 8 2 0 0 0 0 | 395

sci.med 8 2 4 0 8 2 4 0 2 0 4 2 10 325 6 2 4 0 4 8 | 395

sci.space 8 18 14 4 8 6 8 4 0 2 2 0 10 10 353 4 6 0 18 10 | 485

soc.religion.christian 64 2 8 0 0 0 0 2 2 2 4 0 2 22 2 376 8 4 2 112 | 612

talk.politics.guns 10 2 0 0 6 4 6 8 0 2 0 10 0 10 2 4 325 8 178 42 | 617

talk.politics.mideast 16 2 0 0 0 0 2 2 2 2 0 0 2 10 2 2 6 328 8 6 | 390

talk.politics.misc 8 4 18 0 4 0 6 16 8 10 4 8 4 20 22 0 30 42 191 18 | 413

talk.religion.misc 18 0 2 0 0 0 0 0 0 0 0 0 0 0 2 6 4 0 2 102 | 136

-----------------------------------------------------------------

Total 391 471 774 479 487 469 523 440 433 434 411 420 527 467 435 420 403 424 429 400

-----------------------------------------------------------------

Class : alt.atheism

Precision : 0.6317135549872123

Recall : 0.5811764705882353

F-score : 0.6053921568627452

Class : comp.graphics

Precision : 0.6518046709129511

Recall : 0.4561664190193165

F-score : 0.5367132867132867

Class : comp.os.ms-windows.misc

Precision : 0.01808785529715762

Recall : 0.6363636363636364

F-score : 0.03517587939698492

Class : comp.sys.ibm.pc.hardware

Precision : 0.6367432150313153

Recall : 0.3485714285714286

F-score : 0.45051698670605617

Class : comp.sys.mac.hardware

Precision : 0.5811088295687885

Recall : 0.6059957173447538

F-score : 0.5932914046121595

Class : comp.windows.x

Precision : 0.6844349680170576

Recall : 0.5103338632750397

F-score : 0.5846994535519126

Class : misc.forsale

Precision : 0.491395793499044

Recall : 0.8371335504885994

F-score : 0.619277108433735

Class : rec.autos

Precision : 0.8

Recall : 0.7213114754098361

F-score : 0.7586206896551724

Class : rec.motorcycles

Precision : 0.8383371824480369

Recall : 0.8962962962962963

F-score : 0.866348448687351

Class : rec.sport.baseball

Precision : 0.8294930875576036

Recall : 0.9090909090909091

F-score : 0.8674698795180722

Class : rec.sport.hockey

Precision : 0.9416058394160584

Recall : 0.7603143418467584

F-score : 0.8413043478260871

Class : sci.crypt

Precision : 0.8857142857142857

Recall : 0.6220735785953178

F-score : 0.7308447937131631

Class : sci.electronics

Precision : 0.49146110056925996

Recall : 0.6556962025316456

F-score : 0.561822125813449

Class : sci.med

Precision : 0.69593147751606

Recall : 0.8227848101265823

F-score : 0.7540603248259861

Class : sci.space

Precision : 0.8114942528735632

Recall : 0.7278350515463917

F-score : 0.767391304347826

Class : soc.religion.christian

Precision : 0.8952380952380953

Recall : 0.6143790849673203

F-score : 0.7286821705426357

Class : talk.politics.guns

Precision : 0.8064516129032258

Recall : 0.526742301458671

F-score : 0.6372549019607844

Class : talk.politics.mideast

Precision : 0.7735849056603774

Recall : 0.841025641025641

F-score : 0.8058968058968057

Class : talk.politics.misc

Precision : 0.44522144522144524

Recall : 0.46246973365617433

F-score : 0.4536817102137767

Class : talk.religion.misc

Precision : 0.255

Recall : 0.75

F-score : 0.3805970149253732

Total Accuracy:77.36325013276686

**3)**

Classes:

comp.graphics

comp.os.ms-windows.misc

comp.sys.ibm.pc.hardware

comp.sys.mac.hardware

rec.autos

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Confusion Metrics\*\*\*\*\*\*\*\*\*\*\*\*\*\*

comp.graphics comp.os.ms-windows.misc comp.sys.ibm.pc.hardware comp.sys.mac.hardware rec.autos Total

comp.graphics 344 244 30 28 8 | 654

comp.os.ms-windows.misc 0 4 2 2 0 | 8

comp.sys.ibm.pc.hardware 38 398 325 78 6 | 845

comp.sys.mac.hardware 42 110 92 319 0 | 563

rec.autos 10 28 10 24 389 | 461

-----------------------------------------------------------------

Total 434 784 459 451 403

-----------------------------------------------------------------

Class : comp.graphics

Precision : 0.7926267281105991

Recall : 0.5259938837920489

F-score : 0.6323529411764706

Class : comp.os.ms-windows.misc

Precision : 0.00510204081632653

Recall : 0.5

F-score : 0.010101010101010102

Class : comp.sys.ibm.pc.hardware

Precision : 0.7080610021786492

Recall : 0.38461538461538464

F-score : 0.4984662576687116

Class : comp.sys.mac.hardware

Precision : 0.7073170731707317

Recall : 0.566607460035524

F-score : 0.6291913214990138

Class : rec.autos

Precision : 0.9652605459057072

Recall : 0.8438177874186551

F-score : 0.9004629629629631

Total Accuracy:70.60327198364008