1. **Class priors**

Class priors have been calculated as

P(omega = 1) = 0.0426

P(omega = 2) = 0.0516

P(omega = 3) = 0.0508

P(omega = 4) = 0.0521

P(omega = 5) = 0.0510

P(omega = 6) = 0.0525

P(omega = 7) = 0.0516

P(omega = 8) = 0.0525

P(omega = 9) = 0.0529

P(omega = 10) = 0.0527

P(omega = 11) = 0.0531

P(omega = 12) = 0.0527

P(omega = 13) = 0.0524

P(omega = 14) = 0.0527

P(omega = 15) = 0.0526

P(omega = 16) = 0.0532

P(omega = 17) = 0.0484

P(omega = 18) = 0.0500

P(omega = 19) = 0.0412

P(omega = 20) = 0.0334

**2.Results based on Bayesian estimator**

**2.1. Training Data on Bayesian estimator**

Predicted class has been assigned as

where , is the total number of vocabulary, is the number of times the word, , appears in all the documents in that class and n is the total number of words in the documents of class .

**Overall accuracy** for train data = 0.9413

**Class accuracy** for train data:

Group 01 = 0.9667

Group 02 = 0.9208

Group 03 = 0.8794

Group 04 = 0.9302

Group 05 = 0.9409

Group 06 = 0.9493

Group 07 = 0.7732

Group 08 = 0.9662

Group 09 = 0.9631

Group 10 = 0.9714

Group 11 = 0.9783

Group 12 = 0.9798

Group 13 = 0.9239

Group 14 = 0.9764

Group 15 = 0.9798

Group 16 = 0.9833

Group 17 = 0.9853

Group 18 = 0.9699

Group 19 = 0.9698

Group 20 = 0.7633

**Confusion matrix for train data with BE is following-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 464 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 11 | 0 | 1 | 1 | 2 |
| 1 | 535 | 6 | 14 | 1 | 9 | 2 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 2 | 4 | 0 | 0 | 2 | 0 |
| 1 | 10 | 503 | 24 | 1 | 19 | 2 | 0 | 0 | 0 | 0 | 7 | 1 | 1 | 0 | 2 | 0 | 0 | 1 | 0 |
| 0 | 10 | 4 | 546 | 4 | 4 | 6 | 2 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 2 | 0 | 2 | 2 | 1 |
| 2 | 5 | 2 | 7 | 541 | 3 | 1 | 0 | 2 | 0 | 0 | 2 | 1 | 2 | 2 | 3 | 0 | 1 | 1 | 0 |
| 0 | 12 | 7 | 1 | 1 | 562 | 1 | 0 | 1 | 1 | 0 | 2 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| 1 | 3 | 2 | 34 | 7 | 2 | 450 | 18 | 1 | 3 | 3 | 16 | 14 | 5 | 4 | 6 | 5 | 1 | 7 | 0 |
| 1 | 0 | 0 | 3 | 1 | 2 | 3 | 572 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 3 | 1 |
| 0 | 0 | 0 | 1 | 1 | 0 | 5 | 1 | 574 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 6 | 1 | 3 | 0 |
| 0 | 3 | 0 | 1 | 0 | 1 | 1 | 3 | 0 | 577 | 4 | 0 | 0 | 1 | 0 | 1 | 2 | 0 | 0 | 0 |
| 1 | 0 | 1 | 2 | 0 | 1 | 0 | 2 | 0 | 0 | 585 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 2 | 0 |
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 582 | 0 | 1 | 0 | 0 | 3 | 1 | 5 | 0 |
| 0 | 4 | 0 | 15 | 5 | 0 | 3 | 2 | 0 | 0 | 1 | 5 | 546 | 2 | 2 | 2 | 2 | 0 | 2 | 0 |
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 2 | 580 | 0 | 5 | 2 | 0 | 1 | 0 |
| 1 | 2 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 581 | 2 | 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 589 | 2 | 3 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 537 | 2 | 3 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 2 | 0 | 6 | 0 | 547 | 4 | 0 |
| 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 0 | 1 | 0 | 1 | 2 | 2 | 450 | 0 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 40 | 15 | 4 | 4 | 287 |

**2.2 Test Data on Bayesian estimator**

**Overall accuracy** for test data = 0.7812

**Class accuracy** for test data:

Group 01 = 0.7327

Group 02 = 0.7686

Group 03 = 0.5294

Group 04 = 0.7781

Group 05 = 0.7154

Group 06 = 0.7821

Group 07 = 0.5969

Group 08 = 0.9013

Group 09 = 0.8892

Group 10 = 0.8690

Group 11 = 0.9549

Group 12 = 0.9139

Group 13 = 0.6565

Group 14 = 0.8219

Group 15 = 0.8571

Group 16 = 0.9472

Group 17 = 0.8901

Group 18 = 0.8670

Group 19 = 0.5935

Group 20 = 0.3506

**Confusion matrix for test data with BE is following-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 233 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 3 | 46 | 3 | 11 | 7 | 9 |
| 3 | 299 | 5 | 11 | 6 | 22 | 1 | 3 | 2 | 0 | 0 | 18 | 4 | 3 | 7 | 4 | 0 | 0 | 1 | 0 |
| 3 | 33 | 207 | 58 | 11 | 31 | 0 | 1 | 2 | 2 | 1 | 18 | 1 | 4 | 4 | 6 | 0 | 0 | 8 | 1 |
| 0 | 8 | 15 | 305 | 21 | 2 | 4 | 6 | 0 | 0 | 1 | 6 | 23 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 10 | 9 | 37 | 274 | 2 | 4 | 5 | 1 | 1 | 0 | 6 | 16 | 7 | 2 | 0 | 3 | 0 | 6 | 0 |
| 0 | 44 | 7 | 9 | 2 | 305 | 1 | 0 | 2 | 1 | 0 | 10 | 0 | 0 | 3 | 2 | 1 | 1 | 2 | 0 |
| 0 | 8 | 4 | 48 | 21 | 1 | 228 | 33 | 5 | 0 | 1 | 3 | 10 | 2 | 3 | 4 | 2 | 3 | 6 | 0 |
| 0 | 1 | 0 | 2 | 0 | 1 | 5 | 356 | 5 | 2 | 0 | 1 | 4 | 0 | 2 | 1 | 4 | 2 | 9 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 26 | 353 | 2 | 0 | 1 | 1 | 1 | 0 | 1 | 4 | 2 | 5 | 0 |
| 4 | 0 | 0 | 1 | 1 | 3 | 3 | 3 | 1 | 345 | 17 | 2 | 2 | 0 | 0 | 3 | 1 | 2 | 9 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 4 | 381 | 1 | 0 | 2 | 1 | 2 | 0 | 1 | 3 | 0 |
| 0 | 5 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 361 | 3 | 2 | 0 | 2 | 7 | 0 | 8 | 1 |
| 2 | 18 | 0 | 27 | 8 | 3 | 1 | 11 | 2 | 0 | 0 | 46 | 258 | 6 | 3 | 6 | 0 | 2 | 0 | 0 |
| 10 | 7 | 1 | 3 | 0 | 0 | 0 | 4 | 0 | 1 | 0 | 1 | 3 | 323 | 4 | 17 | 3 | 7 | 9 | 0 |
| 3 | 7 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 1 | 4 | 4 | 3 | 336 | 5 | 1 | 2 | 22 | 1 |
| 7 | 3 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 377 | 2 | 2 | 1 | 1 |
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 2 | 0 | 3 | 0 | 1 | 2 | 3 | 324 | 3 | 16 | 4 |
| 9 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 4 | 0 | 0 | 0 | 9 | 4 | 326 | 18 | 0 |
| 6 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 3 | 0 | 3 | 7 | 3 | 95 | 5 | 184 | 1 |
| 47 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 3 | 5 | 71 | 19 | 5 | 8 | 88 |

**3 Results based on Maximum Likelihood estimator**

Predicted class has been assigned as

where , is the number of times the word, , appears in all the documents in that class and is the total number of words in the documents of class .

**3.1). Training Data of Maximum Likelihood estimator**

**Overall accuracy** for train data = 0.9896

**Class accuracy** for train data:

Group 01 = 0.9958

Group 02 = 0.9793

Group 03 = 0.9895

Group 04 = 0.9864

Group 05 = 0.9878

Group 06 = 0.9831

Group 07 = 0.9914

Group 08 = 0.9899

Group 09 = 0.9950

Group 10 = 0.9916

Group 11 = 0.9883

Group 12 = 0.9983

Group 13 = 0.9882

Group 14 = 0.9949

Group 15 = 0.9949

Group 16 = 0.9866

Group 17 = 0.9945

Group 18 = 0.9911

Group 19 = 0.9849

Group 20 = 0.9761

**Confusion matrix for train data-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 478 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 569 | 2 | 4 | 1 | 1 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 4 | 566 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 2 | 1 | 579 | 2 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 568 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 5 | 2 | 0 | 0 | 582 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1 | 577 | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 1 | 586 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 593 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 589 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 1 | 591 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 593 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 584 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 591 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 590 | 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 591 | 2 | 0 | 0 | 3 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 542 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 559 | 1 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 457 | 2 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 1 | 0 | 1 | 367 |

**3.2). Test Data of Maximum Likelihood estimator**

**Overall accuracy** for test data = 0.0946

**Class accuracy** for test data:

Group 01 = 0.9937

Group 02 = 0.0694

Group 03 = 0.0486

Group 04 = 0.0714

Group 05 = 0.0574

Group 06 = 0.0821

Group 07 = 0.1178

Group 08 = 0.0481

Group 09 = 0.0504

Group 10 = 0.0529

Group 11 = 0.0902

Group 12 = 0.0430

Group 13 = 0.0229

Group 14 = 0.0356

Group 15 = 0.0434

Group 16 = 0.0678

Group 17 = 0.0330

Group 18 = 0.0399

Group 19 = 0.0258

Group 20 = 0.0239

**Confusion matrix for test data-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 316 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 351 | 27 | 2 | 1 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| 353 | 3 | 19 | 7 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 1 | 0 | 0 | 0 |
| 350 | 3 | 5 | 28 | 2 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 351 | 2 | 0 | 2 | 22 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 352 | 2 | 0 | 1 | 1 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 318 | 4 | 1 | 4 | 1 | 0 | 45 | 2 | 3 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| 369 | 0 | 0 | 0 | 0 | 0 | 1 | 19 | 2 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 376 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 372 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 21 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 363 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 36 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 375 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 373 | 1 | 0 | 0 | 0 | 1 | 3 | 1 | 0 | 0 | 0 | 3 | 9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 375 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 14 | 0 | 0 | 1 | 0 | 0 | 0 |
| 375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 0 | 0 | 0 |
| 367 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 27 | 0 | 1 | 1 | 1 |
| 350 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 1 | 0 | 1 |
| 358 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 15 | 1 | 0 |
| 297 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 0 | 8 | 0 |
| 237 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 4 | 0 | 0 | 2 | 6 |

**Comment:**

Overall accuracy as well as group specific accuracy for the training data are higher that test data. It is found that MLE overall train accuracy (0.9896) is higher than the Bayes estimation overall train accuracy (0.9413). however, MLE overall test accuracy (0.0946) is much lower that the Bayes estimation overall test accuracy (0.7812) that indicates that MLE estimation overestimates on the training data. As a result, all groups accuracy for MLE is very poor except reference group (group 1). This may happen because for MLE the conditional probability values are zero, and for any new words it considers there is no chance to include in any group except reference group. However, for the Bayes estimation, they have some very small prior probability. Thus, for this problem, Bayesian estimator performs better than MLE estimator.