// this is pre- adding bigrams for count (but the bigrams for count weren’t different than the other results really)

SWITCHBOARD

Naive Bayes

Count:

precision recall f1-score support

f 0.66 0.57 0.61 25907

m 0.51 0.60 0.55 19093

avg / total 0.60 0.59 0.59 45000

Word TF-IDF:

precision recall f1-score support

f 0.64 0.57 0.60 25176

m 0.52 0.59 0.55 19824

avg / total 0.59 0.58 0.58 45000

N-Gram TF-IDF:

precision recall f1-score support

f 0.73 0.56 0.63 29659

m 0.42 0.61 0.49 15341

avg / total 0.62 0.57 0.59 45000

=======================================

Logistic Regression

Count:

precision recall f1-score support

f 0.61 0.58 0.60 23716

m 0.56 0.59 0.58 21284

avg / total 0.59 0.59 0.59 45000

Word TF-IDF:

precision recall f1-score support

f 0.59 0.59 0.59 22428

m 0.59 0.59 0.59 22572

avg / total 0.59 0.59 0.59 45000

N-Gram TF-IDF:

precision recall f1-score support

f 0.47 0.59 0.52 17757

m 0.68 0.56 0.62 27243

avg / total 0.60 0.57 0.58 45000

=======================================

SGD

Count:

precision recall f1-score support

f 0.61 0.58 0.60 23754

m 0.56 0.59 0.57 21246

avg / total 0.59 0.58 0.58 45000

Word TF-IDF:

precision recall f1-score support

f 0.52 0.59 0.56 19999

m 0.64 0.57 0.60 25001

avg / total 0.59 0.58 0.58 45000

N-Gram TF-IDF:

precision recall f1-score support

f 0.16 0.60 0.25 5799

m 0.90 0.52 0.65 39201

avg / total 0.80 0.53 0.60 45000

=======================================

SGD - log loss

Count:

precision recall f1-score support

f 0.59 0.59 0.59 22535

m 0.59 0.59 0.59 22465

avg / total 0.59 0.59 0.59 45000

Word TF-IDF:

precision recall f1-score support

f 0.55 0.59 0.57 21226

m 0.61 0.58 0.59 23774

avg / total 0.58 0.58 0.58 45000

N-Gram TF-IDF:

precision recall f1-score support

f 0.40 0.60 0.48 14924

m 0.74 0.55 0.63 30076

avg / total 0.62 0.57 0.58 45000

MOVIE

Naive Bayes

Count:

precision recall f1-score support

f 0.68 0.59 0.63 19064

m 0.51 0.61 0.56 13436

avg / total 0.61 0.60 0.60 32500

Word TF-IDF:

precision recall f1-score support

f 0.67 0.59 0.63 18735

m 0.52 0.61 0.56 13765

avg / total 0.61 0.60 0.60 32500

N-Gram TF-IDF:

precision recall f1-score support

f 0.59 0.56 0.57 17198

m 0.53 0.56 0.54 15302

avg / total 0.56 0.56 0.56 32500

=======================================

Logistic Regression

Count:

precision recall f1-score support

f 0.64 0.59 0.62 17927

m 0.54 0.60 0.57 14573

avg / total 0.60 0.59 0.60 32500

Word TF-IDF:

precision recall f1-score support

f 0.61 0.59 0.60 16840

m 0.58 0.60 0.59 15660

avg / total 0.60 0.60 0.60 32500

N-Gram TF-IDF:

precision recall f1-score support

f 0.49 0.58 0.53 13812

m 0.64 0.55 0.59 18688

avg / total 0.57 0.56 0.57 32500

=======================================

SGD

Count:

precision recall f1-score support

f 0.68 0.58 0.63 19307

m 0.49 0.61 0.54 13193

avg / total 0.61 0.59 0.59 32500

Word TF-IDF:

precision recall f1-score support

f 0.56 0.59 0.58 15390

m 0.61 0.58 0.59 17110

avg / total 0.59 0.59 0.59 32500

N-Gram TF-IDF:

precision recall f1-score support

f 0.10 0.61 0.18 2722

m 0.93 0.51 0.66 29778

avg / total 0.87 0.52 0.62 32500

=======================================

SGD - log loss

Count:

precision recall f1-score support

f 0.63 0.59 0.61 17530

m 0.55 0.60 0.57 14970

avg / total 0.59 0.59 0.59 32500

Word TF-IDF:

precision recall f1-score support

f 0.58 0.59 0.58 15919

m 0.60 0.58 0.59 16581

avg / total 0.59 0.59 0.59 32500

N-Gram TF-IDF:

precision recall f1-score support

f 0.42 0.58 0.49 11627

m 0.70 0.54 0.61 20873

avg / total 0.60 0.56 0.57 32500

TRAIN SWITCH, TEST MOVIES

Naive Bayes

Count:

precision recall f1-score support

f 0.62 0.53 0.57 76363

m 0.45 0.54 0.49 53637

avg / total 0.55 0.54 0.54 130000

Word TF-IDF:

precision recall f1-score support

f 0.58 0.53 0.55 70347

m 0.49 0.54 0.52 59653

avg / total 0.54 0.54 0.54 130000

N-Gram TF-IDF:

precision recall f1-score support

f 0.64 0.51 0.57 81018

m 0.39 0.52 0.45 48982

avg / total 0.55 0.52 0.52 130000

=======================================

Logistic Regression

Count:

precision recall f1-score support

f 0.56 0.53 0.55 68537

m 0.51 0.54 0.52 61463

avg / total 0.53 0.53 0.53 130000

Word TF-IDF:

precision recall f1-score support

f 0.53 0.54 0.53 64351

m 0.54 0.54 0.54 65649

avg / total 0.54 0.54 0.54 130000

N-Gram TF-IDF:

precision recall f1-score support

f 0.43 0.52 0.47 53550

m 0.61 0.52 0.56 76450

avg / total 0.53 0.52 0.52 130000

=======================================

SGD

Count:

precision recall f1-score support

f 0.60 0.53 0.56 73117

m 0.47 0.54 0.50 56883

avg / total 0.54 0.53 0.54 130000

Word TF-IDF:

precision recall f1-score support

f 0.41 0.55 0.47 48897

m 0.66 0.53 0.59 81103

avg / total 0.57 0.54 0.54 130000

N-Gram TF-IDF:

precision recall f1-score support

f 0.02 0.57 0.04 2257

m 0.99 0.50 0.66 127743

avg / total 0.97 0.50 0.65 130000

=======================================

SGD - log loss

Count:

precision recall f1-score support

f 0.54 0.54 0.54 65742

m 0.53 0.54 0.53 64258

avg / total 0.54 0.54 0.54 130000

Word TF-IDF:

precision recall f1-score support

f 0.47 0.54 0.50 56081

m 0.61 0.53 0.57 73919

avg / total 0.55 0.54 0.54 130000

N-Gram TF-IDF:

precision recall f1-score support

f 0.27 0.54 0.36 32108

m 0.77 0.51 0.62 97892

avg / total 0.65 0.52 0.55 130000

TRAIN MOVIES, TEST SWITCHBOARD

Naive Bayes

Count:

precision recall f1-score support

f 0.50 0.54 0.52 84115

m 0.57 0.53 0.55 95885

avg / total 0.54 0.54 0.54 180000

Word TF-IDF:

precision recall f1-score support

f 0.52 0.54 0.53 86632

m 0.55 0.53 0.54 93368

avg / total 0.54 0.54 0.54 180000

N-Gram TF-IDF:

precision recall f1-score support

f 0.72 0.52 0.60 125786

m 0.33 0.54 0.41 54214

avg / total 0.61 0.53 0.55 180000

=======================================

Logistic Regression

Count:

precision recall f1-score support

f 0.50 0.54 0.52 83675

m 0.57 0.53 0.55 96325

avg / total 0.54 0.53 0.53 180000

Word TF-IDF:

precision recall f1-score support

f 0.47 0.54 0.50 78029

m 0.60 0.53 0.56 101971

avg / total 0.54 0.54 0.54 180000

N-Gram TF-IDF:

precision recall f1-score support

f 0.41 0.54 0.46 68435

m 0.65 0.52 0.58 111565

avg / total 0.56 0.53 0.54 180000

=======================================

SGD

Count:

precision recall f1-score support

f 0.56 0.54 0.55 93807

m 0.52 0.54 0.53 86193

avg / total 0.54 0.54 0.54 180000

Word TF-IDF:

precision recall f1-score support

f 0.41 0.55 0.47 67802

m 0.66 0.53 0.59 112198

avg / total 0.57 0.54 0.54 180000

N-Gram TF-IDF:

precision recall f1-score support

f 0.12 0.56 0.20 19680

m 0.90 0.51 0.65 160320

avg / total 0.82 0.51 0.60 180000

=======================================

SGD - log loss

Count:

precision recall f1-score support

f 0.49 0.55 0.51 80068

m 0.60 0.54 0.56 99932

avg / total 0.55 0.54 0.54 180000

Word TF-IDF:

precision recall f1-score support

f 0.44 0.55 0.49 71845

m 0.64 0.53 0.58 108155

avg / total 0.56 0.54 0.54 180000

N-Gram TF-IDF:

precision recall f1-score support

f 0.32 0.54 0.40 52372

m 0.73 0.52 0.61 127628

avg / total 0.61 0.53 0.55 180000