## Experiment 1

Results after cleaning the larger files (events-manual-examples-1.csv and tagged-irrelevant-1.csv)

This is the result for test data. Cross validation is not applied.

### MODEL: Multinomial Naive Bayes

The precision for this classifier is 0.853115530497

The recall for this classifier is 0.923642533937

The f1 for this classifier is 0.886979275459

The accuracy for this classifier is 0.923642533937

Here is the classification report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Categories** | **precision** | **recall** | **f1-score** | **support** |
|  |  |  |  |  |
| Completion of education | 0.00 | 0.00 | 0.00 | 16 |
| Death of family member | 0.00 | 0.00 | 0.00 | 27 |
| Divorce | 0.00 | 0.00 | 0.00 | 9 |
| Engagement | 0.00 | 0.00 | 0.00 | 6 |
| Holiday | 0.00 | 0.00 | 0.00 | 2 |
| Major purchase - Car | 0.00 | 0.00 | 0.00 | 13 |
| Major purchase - new car | 0.00 | 0.00 | 0.00 | 8 |
| Marriage | 0.00 | 0.00 | 0.00 | 3 |
| New Job | 0.00 | 0.00 | 0.00 | 1 |
| New child | 0.00 | 0.00 | 0.00 | 13 |
| New home purchase | 0.00 | 0.00 | 0.00 | 5 |
| New job | 0.00 | 0.00 | 0.00 | 9 |
| Promotion | 0.00 | 0.00 | 0.00 | 19 |
| Retirement | 0.00 | 0.00 | 0.00 | 4 |
| irrelevant | 0.92 | 1.00 | 0.96 | 1633 |
|  |  |  |  |  |
| **avg / total** | **0.85** | **0.92** | **0.89** | **1768** |

### 3 Fold CV Accuracy:

[0.92945454545454542, 0.9155749636098981, 0.92721979621542938]

The average accuracy for this classifier is 0.92408310176.

## Experiment 2

Results of data set named "all-irrelevant-tagged-samples" and "all-tagged-samples.csv"

All irrelevant are termed individually

### MODEL: Multinomial Naive Bayes

The precision for this classifier is 0.713185069427

The recall for this classifier is 0.756474103586

The f1 for this classifier is 0.69768810976

The accuracy for this classifier is 0.756474103586

Here is the classification report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Categories** | **precision** | **recall** | **f1-score** | **support** |
|  |  |  |  |  |
| completion of education | 0.00 | 0.00 | 0.00 | 21 |
| death of family member | 0.00 | 0.00 | 0.00 | 30 |
| divorce | 0.00 | 0.00 | 0.00 | 6 |
| engagement | 0.00 | 0.00 | 0.00 | 30 |
| holiday | 0.00 | 0.00 | 0.00 | 2 |
| irrelevant - divorce | 0.98 | 0.63 | 0.77 | 232 |
| irrelevant - major purchase - car | 1.00 | 0.05 | 0.10 | 91 |
| irrelevant - marriage | 0.65 | 0.99 | 0.79 | 445 |
| irrelevant - new child/dependent | 0.90 | 0.75 | 0.82 | 217 |
| irrelevant - new home purchase | 0.69 | 0.97 | 0.80 | 438 |
| irrelevant - new job | 0.89 | 0.93 | 0.91 | 367 |
| major purchase - new car | 0.00 | 0.00 | 0.00 | 23 |
| marriage | 0.00 | 0.00 | 0.00 | 36 |
| new child | 0.00 | 0.00 | 0.00 | 13 |
| new home purchase | 0.00 | 0.00 | 0.00 | 17 |
| new job | 0.00 | 0.00 | 0.00 | 20 |
| promotion | 0.00 | 0.00 | 0.00 | 16 |
| retirement | 0.00 | 0.00 | 0.00 | 4 |
|  |  |  |  |  |
| **avg / total** | **0.71** | **0.76** | **0.70** | **2008** |

## Experiment 3

Results of data set named "all-irrelevant-tagged-samples" and "all-tagged-samples.csv"

All irrelevant are grouped as irrelevant

### MODEL: Multinomial Naive Bayes

The precision for this classifier is 0.794655045158

The recall for this classifier is 0.891434262948

The f1 for this classifier is 0.840267156755

The accuracy for this classifier is 0.891434262948

Here is the classification report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Categories** | **precision** | **recall** | **f1-score** | **support** |
|  |  |  |  |  |
| completion of education | 0.00 | 0.00 | 0.00 | 21 |
| death of family member | 0.00 | 0.00 | 0.00 | 30 |
| divorce | 0.00 | 0.00 | 0.00 | 6 |
| engagement | 0.00 | 0.00 | 0.00 | 30 |
| holiday | 0.00 | 0.00 | 0.00 | 2 |
| irrelevant | 0.89 | 1.00 | 0.94 | 1790 |
| major purchase - new car | 0.00 | 0.00 | 0.00 | 23 |
| marriage | 0.00 | 0.00 | 0.00 | 36 |
| new child | 0.00 | 0.00 | 0.00 | 13 |
| new home purchase | 0.00 | 0.00 | 0.00 | 17 |
| new job | 0.00 | 0.00 | 0.00 | 20 |
| promotion | 0.00 | 0.00 | 0.00 | 16 |
| retirement | 0.00 | 0.00 | 0.00 | 4 |
|  |  |  |  |  |
| **avg / total** | **0.79** | **0.89** | **0.84** | **2008** |

### 2 Fold CV Result

[0.89116517285531371, 0.89538855678906915]

The average accuracy for this classifier is 0.893276864822