**Assignment 4: MNB and SVM for Causal Language Detection**By Shivani Sanjay Mahaddalkar

The analysis involves classifying the relationships in the sentences into different categories.  
0: No relationship   
1: Direct Causal  
2: Conditional Causal  
3: Correlational

**Model 1:** Multinomial Naïve Bayes

Using 5-fold cross validation, we compare the average scores of different vectorizers. We pick the vectorizer with the highest average score. Boolean count vectorizer is used which removes stop words. Using that vectorizer we train the model using 60% of the data and test on the remaining 40% of the data.

Accuracy:0.67

The top ten features for each classes are:

|  |  |  |  |
| --- | --- | --- | --- |
| No relationship | Direct Causal | Conditional Causal | Correlational |
| Treatment | Did | Suggest | Cancer |
| Trial | Effect | Disease | Diabetes |
| Findings | Improved | Reduce | Women |
| High | Cancer | Results | Highest |
| Needed | Effective | Cancer | Levels |
| Clinical | Study | Effective | Study |
| Risk | Treatment | Study | Increased |
| Study | Weight | Treatment | Patients |
| Studies | Risk | Weight | Risk |
| Patients | Patients | risk | associated |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix: | Predicted 0 | Predicted 1 | Predicted 2 | Predicted 3 |
| Actual 0 | 477 | 15 | 1 | 79 |
| Actual 1 | 71 | 49 | 1 | 68 |
| Actual 2 | 45 | 6 | 0 | 25 |
| Actual 3 | 48 | 8 | 0 | 332 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1-score |
| 0 | 0.74 | 0.83 | 0.79 |
| 1 | 0.63 | 0.26 | 0.37 |
| 2 | 0.00 | 0.00 | 0.00 |
| 3 | 0.66 | 0.86 | 0.74 |

Error Analysis:

False direct causal predictions:  
Condom use did increase over time in both groups.  
There was no clear improvement in neurological outcomes after TTM in patients with moderate or severe TBI.  
Incident PD patients treated with low GDP solution have less severe systemic inflammation but trends of less ultrafiltration, and more fluid accumulation.

**Model 2:** Support Vector Machine

Using 5-fold cross validation, we compare the average scores of different vectorizers. We pick the vectorizer with the highest average score. Boolean vectorizer is used which removes stop words. Using that vectorizer we train the model using 60% of the data and test on the remaining 40% of the data.

Accuracy: 0.708

The top ten features for false and true reviews are:

|  |  |  |  |
| --- | --- | --- | --- |
| No relationship | Direct Causal | Conditional Causal | Correlational |
| Implications | Contributed | Play role | Variable |
| Appropriate | Does | Appears | Associations |
| Warranted | Benefits | Useful | Increased |
| Required | Did | Protective | Higher |
| Need | Improved | Play | Predict |
| Necessary | Effect | Mediated bmi | Predictor |
| Research | Effects | Relations | Correlated |
| Safety | Resulted | Relations mediated | Association |
| Needed | Effective | Weight | Related |
| studies | Improved | Improve | associated |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix: | Predicted 0 | Predicted 1 | Predicted 2 | Predicted 3 |
| Actual 0 | 515 | 13 | 2 | 42 |
| Actual 1 | 76 | 73 | 5 | 35 |
| Actual 2 | 46 | 9 | 4 | 17 |
| Actual 3 | 70 | 6 | 1 | 311 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1-score |
| 0 | 0.79 | 0.79 | 0.79 |
| 1 | 0.53 | 0.38 | 0.44 |
| 2 | 0.60 | 0.04 | 0.07 |
| 3 | 0.66 | 0.88 | 0.76 |

False direct causal predictions:  
Condom use did increase over time in both groups.  
Patients with body mass index >40\\xa0kg/m2 have greater than twice the risk for complications with odds ratios increasing with increasing body mass index class.  
Dietary cholesterol intake did not have an association with LDL-C level or with risk for coronary artery calcification in apparently healthy Korean adults.

Results:  
Both the models seem to give similar results with not a lot of information given. However, it SVMs perform better in understanding conditional causal, and multinomial naïve bayes seems to misclassify it every time.

For a causal relationship it is very difficult for the models to gauge what would constitute as causal and would require more context as opposed to word tokens. Therefore, it seems to perform with lesser accuracy then simpler tasks like sentiment prediction.