Veterans Affairs WebPage Data Project Report

Team Members:
Aneesh Kalla, Anurag Perakalapudi

10/21/24

Statement of Data Mining Goal:

Veterans Affairs (VA) is a government agency that is supposed to process claims and provide benefits to Veterans. However, Veterans often have a very difficult time actually claiming the benefits they deserve, because the process to go through the VA can be difficult and tedious. The goal of research in this area is to help Veterans with the process of filing for claims. This projects specific goal is to evaluate the quality of embeddings of the datascraped text.

Description of Dataset:

Data was collected from this website: www.knowva.ebenefits.va.gov, a publicly accessible government website about the rules and regulations behind filing for claims with the Veterans. In the website, theres thousands of article webpages that contain these rules. We created our dataset by extracting the text from each of these pages. Each of these articles is also associated with a broader category.

Preprocessing:

The text from these pages was broken into appropriately sized chunks, if the text in an article was over 500 words. Then, each chunk of text was fed through an openAI embedding model to create embeddings for each chunk of text. Duplicate texts were removed, and special non-recognized characters were also deleted.

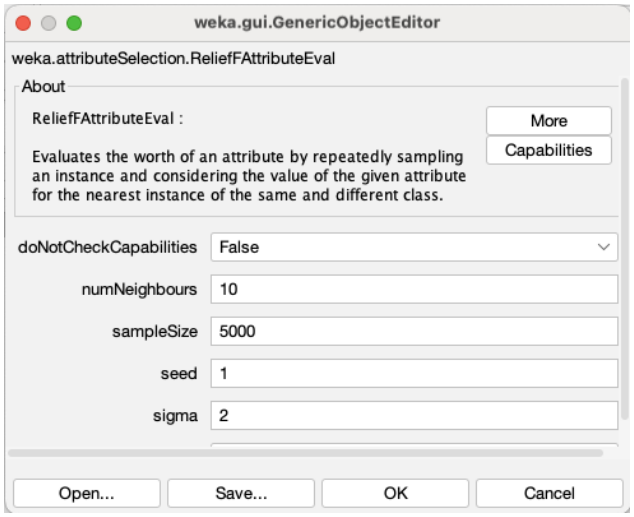
```python
# Function to split text into overlapping chunks
def split_text(content, limit):
    chunks = []
    i = 0
    while i < len(content):
        chunks.append(content[i:i + limit])
        # Move i to the midpoint of the current chunk for overlap
        if i + limit//2 < len(content):  # Ensure not to skip parts of text
            i += limit//2
        else:
            break
    return chunks
```

Attribute Selection Algorithms:

1. **(Non-Weka) No Attribute Removal**
    a. As a baseline method, we keep all the attributes in our data. This is also helpful as a control case to help us put into perspective the other results of our trials.
2. **PrincipalComponents**
    a. PCA (Principal Component Analysis) is a attribute reduction technique that transforms the original attributes into a linear combination of attributes called Principal Components that tries to maximize the variance in the data. The PCA ranks n (# of original attributes) principal components by the amount of variance they capture. By keeping a variance threshold, we can remove some attributes which do not contain enough variance for our liking.
3. **ReliefFAttributeEval**
    a. ReliefF attribute evaluation is an attribute selection algorithm that ranks the attributes by useful they are in distinguishing between the class types.
    ReliefF is effective for handling noisy data since it focuses on the nearest neighbors, which is helpful for our data.
    ReliefF attribute evaluation finds the similarities between different instances and the give a weight for each attribute in relation to the class .Here is an example of an attribute ranking using ReliefFAttributeEval:

```
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 1 class):
        ReliefF Ranking Filter
        Instances sampled: 500
        Number of nearest neighbours (k): 10
        Equal influence nearest neighbours

Ranked attributes:
 0.0427     5 a957
 0.0422     3 a1956
 0.0393     8 a1774
 0.038      9 a2634
 0.0377    16 a1632
 0.0374     4 a1565
 0.0369    11 a2270
 0.0367     7 a2196
 0.036     40 a3016
 0.0358     2 a1999
 0.0349    17 a1679
 0.0346    18 a2202
 0.0344    13 a2662
 0.0343    10 a580
 0.0336    71 a496

Selected attributes: 5,3,8,9,16,4,11,7,40,2,17,18,13,10,71 : 15
```

Here are the hyperparameters we are initializing the algorithm with:



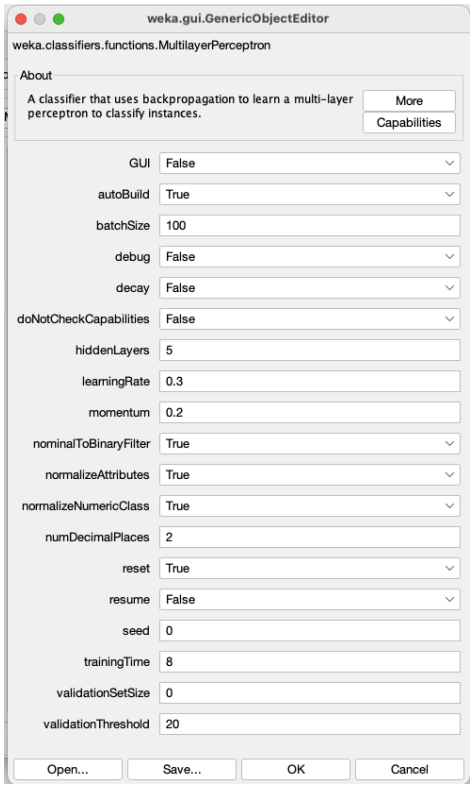
4. **GainRatioAttributeEval**
    a. GainRatioAttribute Eval ranks attributes by their gain ratio in relation to the class attribute. Entropy is calculated by the following formula where $P(C_i)$ is the probability of the attribute predicting the class.

$$-\sum_{i=1}^{c} P(C_i)\log_2 P(C_i)$$

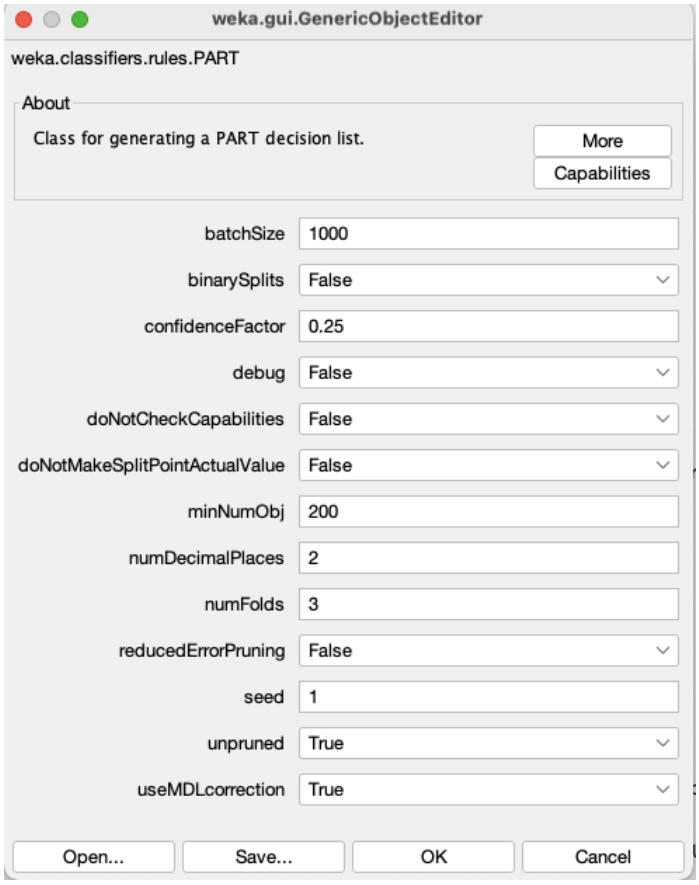
5. **Correlation**
    a. Correlation ranks attributes based on their correlation to each potential output class, and the attributes that have the highest correlation to one or more output classes are chosen

Classification Algorithms:

1. **Multi-layer Perceptron (MLP):**
    a. A Multi-layer Perceptron is a simplistic neural network that takes the instances as the input and the class as the output. We can tune hyperparameters for the MLP. Here are our hyperparameters:



2. **PART (Partial Decision Trees)**
    a. PART (Partial Decision Trees) is a rule-based classification algorithm with a connection with the a decision tree. It works by making a decision tree for the dataset and then extracting rules from the trees and modifying the new trees until PART creates a set of rules for the dataset. Here are the hyperparameters for our PART initialization.



3. **Naive Bayes**
    a. Probabilistic Classifier based on calculating the prior probability based on the formula to the side. P(C|X) describes the function for the previous feature X for the class C.

$$P(C|X) \propto P(C) \prod_{i=1}^{n} P(x_i|C)$$

4. **J48**
    a. J48 is Weka's implementation of the C4.5 algorithm, which works by building a decision tree by recursively splitting a dataset based on which feature has the most significant information gain.

Results:

No Attribute Removal with MLP:

```
=== Summary ===

Correctly Classified Instances       80408               98.5296 %
Incorrectly Classified Instances      1200                1.4704 %
Kappa statistic                          0.9828
Mean absolute error                      0.0075
Root mean squared error                  0.0529
Relative absolute error                  3.9617 %
Root relative squared error             17.1303 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.993    0.002    0.988      0.993    0.991      0.989    0.999     0.998     Outreach Transation and Economic Development
                 0.994    0.003    0.983      0.994    0.988      0.987    1.000     0.998     Home and Auto
                 0.994    0.002    0.989      0.994    0.991      0.990    1.000     0.998     Life Insurance
                 0.993    0.003    0.986      0.993    0.990      0.988    1.000     0.998     Veteran Readiness and Employment
                 0.990    0.003    0.984      0.990    0.987      0.985    0.999     0.997     Fiduciary Program
                 0.000    0.000    0.000      0.000    0.000      -0.000   0.571     0.057     eBenefits
                 0.984    0.004    0.974      0.984    0.979      0.976    0.999     0.995     Education Service
                 0.996    0.001    0.992      0.996    0.994      0.993    1.000     0.999     Stakeholder Enterprise Portal
                 0.000    0.000    ?          0.000    ?          ?        0.577     0.023     Burial and Memorial
Weighted Avg.    0.985    0.002    ?          0.985    ?          ?        0.997     0.991

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 11306     9     7    15    18     0    24     4     0 |     a = Outreach Transation and Economic Development
     2 10770    26    16    11     0    14     0     0 |     b = Home and Auto
     5    41 11023     6     9     0     8     3     0 |     c = Life Insurance
    14    11    13 14353     3     0    40    17     0 |     d = Veteran Readiness and Employment
    22    10    15     5 10856     1    47    15     0 |     e = Fiduciary Program
    57    44    31    73    69     0    94    25     0 |     f = eBenefits
    28    23    11    48    35     0 10637    27     0 |     g = Education Service
     1     0     3    19     8     0    10 11463     0 |     h = Stakeholder Enterprise Portal
     6    47    19    18    24     0    48     1     0 |     i = Burial and Memorial
```

No Attribute Selection with PART:

```
=== Summary ===

Correctly Classified Instances       68893               84.4194 %
Incorrectly Classified Instances     12715               15.5806 %
Kappa statistic                          0.8183
Mean absolute error                      0.0462
Root mean squared error                  0.1589
Relative absolute error                 24.2509 %
Root relative squared error             51.4817 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.882    0.025    0.851      0.882    0.866      0.844    0.978     0.928     Outreach Transation and Economic Development
                 0.823    0.027    0.824      0.823    0.824      0.797    0.970     0.891     Home and Auto
                 0.883    0.028    0.833      0.883    0.857      0.835    0.979     0.926     Life Insurance
                 0.865    0.028    0.869      0.865    0.867      0.838    0.976     0.930     Veteran Readiness and Employment
                 0.840    0.020    0.868      0.840    0.854      0.832    0.974     0.912     Fiduciary Program
                 0.000    0.000    ?          0.000    ?          ?        0.900     0.094     eBenefits
                 0.727    0.036    0.757      0.727    0.741      0.703    0.948     0.813     Education Service
                 0.911    0.013    0.918      0.911    0.915      0.901    0.986     0.958     Stakeholder Enterprise Portal
                 0.613    0.004    0.224      0.613    0.328      0.369    0.926     0.162     Burial and Memorial
Weighted Avg.    0.844    0.025    ?          0.844    ?          ?        0.973     0.905

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 10035   147   112   285   184     0   523    78    19 |     a = Outreach Transation and Economic Development
   204  8917   806   236   217     0   305   123    31 |     b = Home and Auto
   132   529  9794   144   182     0   241    41    32 |     c = Life Insurance
   351   276   180 12498   173     0   730   200    43 |     d = Veteran Readiness and Employment
   358   288   272   269  9212     0   399   135    38 |     e = Fiduciary Program
    43    53    76    32    32     0    52    23    82 |     f = eBenefits
   540   491   349   734   416     0  7858   329    92 |     g = Education Service
   130   103   139   188   187     0   269 10479     9 |     h = Stakeholder Enterprise Portal
     4    13    24     4     8     0     9     1   100 |     i = Burial and Memorial
```

No Attribute Selection with Naive Bayes:

```
=== Summary ===

Correctly Classified Instances       79008               96.814  %
Incorrectly Classified Instances      2600                3.186  %
Kappa statistic                          0.9629
Mean absolute error                      0.0074
Root mean squared error                  0.0797
Relative absolute error                  3.8683 %
Root relative squared error             25.8314 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.961    0.001    0.993      0.961    0.977      0.973    0.996     0.992     Outreach Transation and Economic Development
                 0.970    0.002    0.984      0.970    0.977      0.974    0.997     0.992     Home and Auto
                 0.982    0.005    0.969      0.982    0.976      0.972    0.998     0.989     Life Insurance
                 0.967    0.003    0.984      0.967    0.975      0.970    0.998     0.994     Veteran Readiness and Employment
                 0.961    0.002    0.988      0.961    0.974      0.971    0.998     0.995     Fiduciary Program
                 0.774    0.005    0.415      0.774    0.540      0.564    0.996     0.614     eBenefits
                 0.958    0.008    0.947      0.958    0.952      0.945    0.992     0.984     Education Service
                 0.983    0.000    1.000      0.983    0.991      0.990    1.000     0.999     Stakeholder Enterprise Portal
                 0.994    0.008    0.205      0.994    0.340      0.450    0.999     0.726     Burial and Memorial
Weighted Avg.    0.968    0.003    0.977      0.968    0.972      0.968    0.997     0.990

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 10943    16   104    43    37    58    97     0    85 |     a = Outreach Transation and Economic Development
     2 10513    35    42    23    72    50     1   101 |     b = Home and Auto
     0    57 10896     9     5    50    11     0    67 |     c = Life Insurance
    34    20    95 13974    14    72   178     1    63 |     d = Veteran Readiness and Employment
    23    32     9    23 10546    50   190     0    98 |     e = Fiduciary Program
     0     1     4     0     2   304     2     0    80 |     f = eBenefits
    18    40    27   105    46    96 10359     2   116 |     g = Education Service
     0     3    73    11     3    29    56 11311    18 |     h = Stakeholder Enterprise Portal
     0     0     0     0     0     1     0     0   162 |     i = Burial and Memorial
```

No Attribute Selection with J48:

```
=== Summary ===

Correctly Classified Instances       64075               78.5156 %
Incorrectly Classified Instances     17533               21.4844 %
Kappa statistic                          0.7495
Mean absolute error                      0.0495
Root mean squared error                  0.2107
Relative absolute error                 25.963  %
Root relative squared error             68.2649 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.815    0.034    0.796      0.815    0.805      0.773    0.887     0.687     Outreach Transation and Economic Development
                 0.758    0.039    0.747      0.758    0.752      0.714    0.855     0.605     Home and Auto
                 0.822    0.027    0.829      0.822    0.825      0.798    0.901     0.710     Life Insurance
                 0.822    0.037    0.829      0.822    0.825      0.788    0.890     0.714     Veteran Readiness and Employment
                 0.784    0.032    0.790      0.784    0.787      0.754    0.867     0.653     Fiduciary Program
                 0.244    0.003    0.268      0.244    0.256      0.253    0.645     0.091     eBenefits
                 0.643    0.058    0.629      0.643    0.636      0.580    0.768     0.469     Education Service
                 0.861    0.019    0.883      0.861    0.872      0.852    0.921     0.779     Stakeholder Enterprise Portal
                 0.239    0.001    0.333      0.239    0.279      0.281    0.694     0.114     Burial and Memorial
Weighted Avg.    0.785    0.035    0.786      0.785    0.785      0.751    0.870     0.659

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
  9274   318   162   392   349    25   709   149     5 |     a = Outreach Transation and Economic Development
   333  8214   651   372   386    40   658   162    23 |     b = Home and Auto
   171   671  9116   225   325    25   450   108     4 |     c = Life Insurance
   419   442   190 11876   301    42   928   240    13 |     d = Veteran Readiness and Employment
   423   384   249   318  8601    34   732   225     5 |     e = Fiduciary Program
    42    65    26    44    34    96    56    18    12 |     f = eBenefits
   810   649   450   846   642    48  6949   402    13 |     g = Education Service
   171   229   146   245   231    32   537  9910     3 |     h = Stakeholder Enterprise Portal
     5    23    10    14    24    16    29     3    39 |     i = Burial and Memorial
```

PCA with MLP:

```
=== Summary ===

Correctly Classified Instances       78857               96.629  %
Incorrectly Classified Instances      2751                3.371  %
Kappa statistic                          0.9607
Mean absolute error                      0.0151
Root mean squared error                  0.0789
Relative absolute error                  7.939  %
Root relative squared error             25.573  %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.967    0.003    0.979      0.967    0.973      0.969    0.999     0.995     Outreach Transation and Economic Development
                 0.984    0.009    0.945      0.984    0.964      0.959    0.998     0.983     Home and Auto
                 0.986    0.006    0.962      0.986    0.974      0.970    0.999     0.989     Life Insurance
                 0.964    0.004    0.982      0.964    0.973      0.967    0.997     0.991     Veteran Readiness and Employment
                 0.968    0.004    0.972      0.968    0.970      0.965    0.998     0.992     Fiduciary Program
                 0.000    0.000    0.000      0.000    0.000      -0.000   0.583     0.038     eBenefits
                 0.957    0.011    0.931      0.957    0.944      0.935    0.996     0.982     Education Service
                 0.986    0.002    0.990      0.986    0.988      0.986    1.000     0.999     Stakeholder Enterprise Portal
                 0.000    0.000    ?          0.000    ?          ?        0.540     0.003     Burial and Memorial
Weighted Avg.    0.966    0.005    ?          0.966    ?          ?        0.995     0.984

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 11005    16    80    52    49     0   179     2     0 |     a = Outreach Transation and Economic Development
     5 10669    61    30    34     0    40     0     0 |     b = Home and Auto
     2   106 10943     9     7     0    22     6     0 |     c = Life Insurance
    58   108   110 13931    14     0   180    50     0 |     d = Veteran Readiness and Employment
    34    86     7     8 10621     1   192    22     0 |     e = Fiduciary Program
    47   154    37    36    55     0    58     6     0 |     f = eBenefits
    57    97    79    85   118     0 10340    33     0 |     g = Education Service
     1     2    42    32    20     0    59 11348     0 |     h = Stakeholder Enterprise Portal
    31    57    22     0    14     0    39     0     0 |     i = Burial and Memorial
```

PCA with PART:

```
=== Summary ===

Correctly Classified Instances       77051               94.416  %
Incorrectly Classified Instances      4557                5.584  %
Kappa statistic                          0.9349
Mean absolute error                      0.0179
Root mean squared error                  0.103
Relative absolute error                  9.4161 %
Root relative squared error             33.3763 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.953    0.006    0.961      0.953    0.957      0.950    0.992     0.976     Outreach Transation and Economic Development
                 0.949    0.009    0.941      0.949    0.945      0.937    0.989     0.968     Home and Auto
                 0.976    0.007    0.958      0.976    0.967      0.962    0.996     0.978     Life Insurance
                 0.956    0.010    0.952      0.956    0.954      0.944    0.994     0.978     Veteran Readiness and Employment
                 0.934    0.008    0.946      0.934    0.940      0.931    0.989     0.966     Fiduciary Program
                 0.461    0.002    0.514      0.461    0.486      0.484    0.921     0.369     eBenefits
                 0.871    0.018    0.879      0.871    0.875      0.856    0.971     0.904     Education Service
                 0.987    0.002    0.985      0.987    0.986      0.984    0.998     0.991     Stakeholder Enterprise Portal
                 0.442    0.001    0.450      0.442    0.446      0.445    0.924     0.322     Burial and Memorial
Weighted Avg.    0.944    0.009    0.944      0.944    0.944      0.935    0.990     0.963

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 10850    27    74   107   102     8   206     5     4 |     a = Outreach Transation and Economic Development
    25 10289   117    68    88    36   194    15     7 |     b = Home and Auto
    10   127 10827    33    18    14    48    13     5 |     c = Life Insurance
    84    71    75 13815    28    32   323    14     9 |     d = Veteran Readiness and Employment
    89    89    53    43 10242     8   414    28     5 |     e = Fiduciary Program
    27    16    33    42    19   181    40    10    25 |     f = eBenefits
   187   288    83   362   310    45  9420    83    31 |     g = Education Service
    17    23    24    30     7     3    43 11355     2 |     h = Stakeholder Enterprise Portal
     4     2    16     8     8    25    28     0    72 |     i = Burial and Memorial
```

PCA with Naive Bayes:

```
=== Summary ===

Correctly Classified Instances       78991               96.7932 %
Incorrectly Classified Instances      2617                3.2068 %
Kappa statistic                          0.9626
Mean absolute error                      0.0098
Root mean squared error                  0.0747
Relative absolute error                  5.1397 %
Root relative squared error             24.2103 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.968    0.002    0.989      0.968    0.978      0.975    0.997     0.994     Outreach Transation and Economic Development
                 0.969    0.004    0.977      0.969    0.973      0.969    0.999     0.995     Home and Auto
                 0.987    0.007    0.959      0.987    0.973      0.969    0.999     0.989     Life Insurance
                 0.970    0.005    0.978      0.970    0.974      0.968    0.998     0.995     Veteran Readiness and Employment
                 0.961    0.003    0.982      0.961    0.972      0.967    0.999     0.996     Fiduciary Program
                 0.669    0.004    0.453      0.669    0.540      0.548    0.996     0.585     eBenefits
                 0.943    0.010    0.934      0.943    0.938      0.929    0.995     0.983     Education Service
                 0.987    0.000    0.998      0.987    0.992      0.991    0.999     0.999     Stakeholder Enterprise Portal
                 0.877    0.002    0.417      0.877    0.565      0.604    0.999     0.696     Burial and Memorial
Weighted Avg.    0.968    0.004    0.971      0.968    0.969      0.964    0.998     0.990

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 11021    10   103    43    34    22   137     1    12 |     a = Outreach Transation and Economic Development
     3 10506    54    61    33    70    89     2    21 |     b = Home and Auto
     1    78 10953    10     4    33     6     1     9 |     c = Life Insurance
    42    28   113 14013    12    61   169     9     4 |     d = Veteran Readiness and Employment
    17    34    26    27 10546    14   280     2    25 |     e = Fiduciary Program
     1    18    26     6     7   263     8     1    63 |     f = eBenefits
    62    77    64   158    97    85 10189    11    66 |     g = Education Service
     0     5    82     7     5    18    30 11357     0 |     h = Stakeholder Enterprise Portal
     0     0     3     0     1    15     1     0   143 |     i = Burial and Memorial
```

PCA with J48:

```
=== Summary ===

Correctly Classified Instances       76998               94.351  %
Incorrectly Classified Instances      4610                5.649  %
Kappa statistic                          0.9341
Mean absolute error                      0.0139
Root mean squared error                  0.1084
Relative absolute error                  7.2795 %
Root relative squared error             35.127  %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.956    0.008    0.950      0.956    0.953      0.945    0.975     0.920     Outreach Transation and Economic Development
                 0.947    0.009    0.943      0.947    0.945      0.937    0.969     0.910     Home and Auto
                 0.970    0.005    0.967      0.970    0.969      0.964    0.986     0.944     Life Insurance
                 0.959    0.009    0.960      0.959    0.960      0.951    0.978     0.929     Veteran Readiness and Employment
                 0.932    0.010    0.936      0.932    0.934      0.923    0.964     0.888     Fiduciary Program
                 0.509    0.002    0.525      0.509    0.517      0.515    0.777     0.310     eBenefits
                 0.867    0.019    0.873      0.867    0.870      0.850    0.914     0.784     Education Service
                 0.988    0.002    0.985      0.988    0.987      0.984    0.994     0.979     Stakeholder Enterprise Portal
                 0.442    0.001    0.537      0.442    0.485      0.486    0.813     0.328     Burial and Memorial
Weighted Avg.    0.944    0.009    0.943      0.944    0.943      0.935    0.968     0.905

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 10878    29    39    80   100    15   228    11     3 |     a = Outreach Transation and Economic Development
    26 10261   109    69   108    20   235     7     4 |     b = Home and Auto
    41   112 10763    33    31    16    75    21     3 |     c = Life Insurance
    68    58    63 13864    24    24   328    17     5 |     d = Veteran Readiness and Employment
   147   125    35    27 10221    15   374    23     4 |     e = Fiduciary Program
    18    29    21    34    11   200    49     7    24 |     f = eBenefits
   257   247    73   315   403    42  9368    85    19 |     g = Education Service
     7    12    14    17    17    10    56 11371     0 |     h = Stakeholder Enterprise Portal
     5     4    12     3     7    39    18     3    72 |     i = Burial and Memorial
```

ReliefF with MLP:

```
=== Summary ===

Correctly Classified Instances       70505               86.3947 %
Incorrectly Classified Instances     11103               13.6053 %
Kappa statistic                          0.8413
Mean absolute error                      0.0475
Root mean squared error                  0.1531
Relative absolute error                 24.943  %
Root relative squared error             49.5952 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.887    0.017    0.892      0.887    0.890      0.872    0.988     0.935     Outreach Transation and Economic Development
                 0.868    0.023    0.854      0.868    0.861      0.840    0.980     0.923     Home and Auto
                 0.862    0.019    0.876      0.862    0.869      0.848    0.984     0.926     Life Insurance
                 0.853    0.020    0.901      0.853    0.876      0.851    0.978     0.938     Veteran Readiness and Employment
                 0.896    0.031    0.818      0.896    0.855      0.833    0.981     0.921     Fiduciary Program
                 0.000    0.000    ?          0.000    ?          ?        0.728     0.021     eBenefits
                 0.768    0.034    0.777      0.768    0.773      0.738    0.950     0.823     Education Service
                 0.954    0.014    0.917      0.954    0.935      0.925    0.996     0.975     Stakeholder Enterprise Portal
                 0.000    0.000    ?          0.000    ?          ?        0.644     0.005     Burial and Memorial
Weighted Avg.    0.864    0.022    ?          0.864    ?          ?        0.978     0.915

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 10099   194    44   183   421     0   438     4     0 |     a = Outreach Transation and Economic Development
   157  9410   669   126   290     0    84   103     0 |     b = Home and Auto
    28   759  9561   106   310     0   281    50     0 |     c = Life Insurance
   246   131   144 12326   382     0   998   224     0 |     d = Veteran Readiness and Employment
   291   191   156   151  9829     0   253   100     0 |     e = Fiduciary Program
    77    78    19    56    70     0    74    19     0 |     f = eBenefits
   407   148   242   635   592     0  8302   483     0 |     g = Education Service
     0    75    44    72   108     0   227 10978     0 |     h = Stakeholder Enterprise Portal
    19    32    40    28    13     0    26     5     0 |     i = Burial and Memorial
```

ReliefF with PART:

```
=== Summary ===

Correctly Classified Instances       65090               79.7593 %
Incorrectly Classified Instances     16518               20.2407 %
Kappa statistic                          0.7638
Mean absolute error                      0.0625
Root mean squared error                  0.1883
Relative absolute error                 32.7718 %
Root relative squared error             61.0073 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.849    0.022    0.864      0.849    0.856      0.833    0.958     0.892     Outreach Transation and Economic Development
                 0.787    0.032    0.791      0.787    0.789      0.756    0.943     0.811     Home and Auto
                 0.797    0.036    0.779      0.797    0.788      0.754    0.947     0.816     Life Insurance
                 0.811    0.039    0.818      0.811    0.814      0.774    0.943     0.865     Veteran Readiness and Employment
                 0.803    0.038    0.764      0.803    0.783      0.749    0.940     0.813     Fiduciary Program
                 0.048    0.001    0.176      0.048    0.076      0.090    0.726     0.042     eBenefits
                 0.664    0.049    0.676      0.664    0.670      0.620    0.898     0.670     Education Service
                 0.898    0.019    0.886      0.898    0.892      0.874    0.970     0.919     Stakeholder Enterprise Portal
                 0.025    0.001    0.089      0.025    0.038      0.046    0.730     0.023     Burial and Memorial
Weighted Avg.    0.798    0.033    0.794      0.798    0.796      0.763    0.942     0.824

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
  9668   186   120   399   387     7   533    79     4 |     a = Outreach Transation and Economic Development
   183  8526   938   296   377    21   347   147     4 |     b = Home and Auto
   100   898  8844   223   453     7   470    95     5 |     c = Life Insurance
   354   286   343 11714   422    34   997   293     8 |     d = Veteran Readiness and Employment
   331   308   331   394  8814     6   554   230     3 |     e = Fiduciary Program
    50    81    26    59    60    19    82    16     0 |     f = eBenefits
   460   332   525  1055   771     8  7174   467    17 |     g = Education Service
    41   134   192   159   227     5   419 10327     0 |     h = Stakeholder Enterprise Portal
     9    32    30    30    20     1    36     1     4 |     i = Burial and Memorial
```

ReliefF with Naive Bayes:

```
=== Summary ===

Correctly Classified Instances       71176               87.2169 %
Incorrectly Classified Instances     10432               12.7831 %
Kappa statistic                          0.8509
Mean absolute error                      0.0394
Root mean squared error                  0.1438
Relative absolute error                 20.6506 %
Root relative squared error             46.5732 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.913    0.012    0.926      0.913    0.920      0.907    0.994     0.974     Outreach Transation and Economic Development
                 0.863    0.020    0.871      0.863    0.867      0.847    0.990     0.943     Home and Auto
                 0.876    0.018    0.882      0.876    0.879      0.860    0.990     0.946     Life Insurance
                 0.874    0.030    0.861      0.874    0.867      0.839    0.984     0.946     Veteran Readiness and Employment
                 0.874    0.023    0.853      0.874    0.863      0.842    0.988     0.940     Fiduciary Program
                 0.216    0.002    0.392      0.216    0.279      0.289    0.955     0.243     eBenefits
                 0.781    0.037    0.765      0.781    0.773      0.738    0.970     0.856     Education Service
                 0.948    0.005    0.967      0.948    0.957      0.950    0.998     0.992     Stakeholder Enterprise Portal
                 0.356    0.002    0.315      0.356    0.334      0.333    0.984     0.316     Burial and Memorial
Weighted Avg.    0.872    0.021    0.872      0.872    0.872      0.851    0.988     0.939

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 10392   104   100   240   204    18   309    10     6 |     a = Outreach Transation and Economic Development
    84  9352   624   233   272    24   179    39    32 |     b = Home and Auto
    24   587  9720   128   280     9   296    24    27 |     c = Life Insurance
   155   124   182 12634   241    19  1042    46     8 |     d = Veteran Readiness and Employment
   218   229   142   282  9585    23   414    52    26 |     e = Fiduciary Program
    39    83     7    40    46    85    78    11     4 |     f = eBenefits
   299   154   190   954   512    38  8444   195    23 |     g = Education Service
     7    69    43   143    92     1   243 10906     0 |     h = Stakeholder Enterprise Portal
     1    36     8    23    10     0    27     0    58 |     i = Burial and Memorial
```

ReliefF with J48:

```
=== Summary ===

Correctly Classified Instances       61561               75.435  %
Incorrectly Classified Instances     20047               24.565  %
Kappa statistic                          0.7135
Mean absolute error                      0.0586
Root mean squared error                  0.2227
Relative absolute error                 30.7357 %
Root relative squared error             72.1519 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.833    0.030    0.816      0.833    0.824      0.796    0.904     0.731     Outreach Transation and Economic Development
                 0.737    0.044    0.720      0.737    0.728      0.686    0.847     0.599     Home and Auto
                 0.751    0.041    0.740      0.751    0.746      0.705    0.869     0.620     Life Insurance
                 0.778    0.049    0.775      0.778    0.776      0.728    0.865     0.669     Veteran Readiness and Employment
                 0.742    0.039    0.749      0.742    0.745      0.706    0.858     0.620     Fiduciary Program
                 0.066    0.003    0.084      0.066    0.074      0.071    0.541     0.017     eBenefits
                 0.598    0.059    0.606      0.598    0.602      0.542    0.760     0.449     Education Service
                 0.859    0.019    0.882      0.859    0.870      0.849    0.925     0.786     Stakeholder Enterprise Portal
                 0.098    0.001    0.163      0.098    0.123      0.125    0.584     0.022     Burial and Memorial
Weighted Avg.    0.754    0.040    0.753      0.754    0.754      0.714    0.860     0.638

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
  9478   285   143   394   381    39   603    52     8 |     a = Outreach Transation and Economic Development
   279  7983  1041   429   427    60   430   167    23 |     b = Home and Auto
   160  1063  8331   336   447    25   591   132    10 |     c = Life Insurance
   441   384   392 11246   437    35  1244   260    12 |     d = Veteran Readiness and Employment
   460   484   461   469  8137    37   693   224     6 |     e = Fiduciary Program
    34    88    33    51    55    26    84    19     3 |     f = eBenefits
   662   516   636  1287   707    53  6464   465    19 |     g = Education Service
    91   253   190   284   261    27   517  9880     1 |     h = Stakeholder Enterprise Portal
     9    37    24    23    11     8    33     2    16 |     i = Burial and Memorial
```

GainRatioAttributeEval with MLP:

```
=== Summary ===

Correctly Classified Instances       67907               83.2112 %
Incorrectly Classified Instances     13701               16.7888 %
Kappa statistic                          0.8041
Mean absolute error                      0.0583
Root mean squared error                  0.1684
Relative absolute error                 30.5728 %
Root relative squared error             54.5648 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.854    0.038    0.786      0.854    0.818      0.788    0.978     0.881     Outreach Transation and Economic Development
                 0.858    0.023    0.853      0.858    0.856      0.834    0.979     0.915     Home and Auto
                 0.908    0.023    0.860      0.908    0.883      0.865    0.989     0.925     Life Insurance
                 0.872    0.019    0.910      0.872    0.890      0.868    0.981     0.946     Veteran Readiness and Employment
                 0.902    0.039    0.782      0.902    0.838      0.813    0.979     0.914     Fiduciary Program
                 0.000    0.000    ?          0.000    ?          ?        0.549     0.006     eBenefits
                 0.555    0.038    0.689      0.555    0.615      0.568    0.893     0.645     Education Service
                 0.897    0.016    0.904      0.897    0.900      0.884    0.990     0.943     Stakeholder Enterprise Portal
                 0.000    0.000    ?          0.000    ?          ?        0.781     0.010     Burial and Memorial
Weighted Avg.    0.832    0.027    ?          0.832    ?          ?        0.968     0.879

=== Confusion Matrix ===

    a     b     c     d     e     f     g     h     i   <-- classified as
 9716   175    79   139   414     0   848    12     0 |     a = Outreach Transation and Economic Development
  101  9303   757    40   326     0   171   141     0 |     b = Home and Auto
   18   481 10070    27   341     0   147    11     0 |     c = Life Insurance
  285    87   141 12600   336     0   785   217     0 |     d = Veteran Readiness and Employment
  231   172   193   143  9898     0   218   116     0 |     e = Fiduciary Program
   93    41     8    28    82     0   113    28     0 |     f = eBenefits
 1915   399   329   639   964     0  6002   561     0 |     g = Education Service
    7   235    94   209   292     0   349 10318     0 |     h = Stakeholder Enterprise Portal
    1     7    38    23    10     0    75     9     0 |     i = Burial and Memorial
```

GainRatioAttributeEval with PART:

```
=== Summary ===

Correctly Classified Instances       63198               77.4409 %
Incorrectly Classified Instances     18410               22.5591 %
Kappa statistic                          0.7367
Mean absolute error                      0.0667
Root mean squared error                  0.1956
Relative absolute error                 34.9928 %
Root relative squared error             63.3792 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.777    0.034    0.788      0.777    0.783      0.748    0.943     0.832     Outreach Transation and Economic Development
               0.797    0.036    0.771      0.797    0.784      0.750    0.946     0.820     Home and Auto
               0.840    0.026    0.838      0.840    0.839      0.814    0.960     0.872     Life Insurance
               0.824    0.037    0.828      0.824    0.826      0.789    0.951     0.878     Veteran Readiness and Employment
               0.801    0.038    0.765      0.801    0.783      0.748    0.942     0.822     Fiduciary Program
               0.031    0.001    0.128      0.031    0.049      0.060    0.715     0.035     eBenefits
               0.526    0.065    0.553      0.526    0.539      0.471    0.861     0.533     Education Service
               0.868    0.025    0.850      0.868    0.859      0.836    0.965     0.895     Stakeholder Enterprise Portal
               0.067    0.001    0.153      0.067    0.094      0.100    0.762     0.066     Burial and Memorial
Weighted Avg.  0.774    0.037    0.770      0.774    0.772      0.736    0.938     0.806

=== Confusion Matrix ===

    a     b     c     d     e     f     g     h     i   <-- classified as
 8842   222    91   379   407    18  1310   113     1 |     a = Outreach Transation and Economic Development
  128  8639   659   182   413    13   503   297     5 |     b = Home and Auto
   73   826  9323   133   337     3   305    85    10 |     c = Life Insurance
  305   210   203 11910   313     4  1198   308     0 |     d = Veteran Readiness and Employment
  287   346   348   308  8789     4   618   269     2 |     e = Fiduciary Program
   57    64    11    49    69    12   102    26     3 |     f = eBenefits
 1459   560   347  1165   885    32  5684   653    24 |     g = Education Service
   62   300   129   232   263     8   506  9988    16 |     h = Stakeholder Enterprise Portal
    3    31    14    29    13     0    55     7    11 |     i = Burial and Memorial
```

GainRatioAttributeEval with Naive Bayes:

```
=== Summary ===

Correctly Classified Instances       69357               84.988  %
Incorrectly Classified Instances     12251               15.012  %
Kappa statistic                          0.825
Mean absolute error                      0.0442
Root mean squared error                  0.1548
Relative absolute error                 23.1724 %
Root relative squared error             50.1387 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.844    0.025    0.846      0.844    0.845      0.820    0.983     0.927     Outreach Transation and Economic Development
               0.868    0.020    0.870      0.868    0.869      0.849    0.989     0.943     Home and Auto
               0.920    0.014    0.911      0.920    0.916      0.902    0.994     0.956     Life Insurance
               0.894    0.025    0.886      0.894    0.890      0.866    0.986     0.958     Veteran Readiness and Employment
               0.862    0.022    0.858      0.862    0.860      0.838    0.987     0.937     Fiduciary Program
               0.153    0.002    0.273      0.153    0.196      0.201    0.958     0.170     eBenefits
               0.658    0.052    0.657      0.658    0.658      0.605    0.941     0.725     Education Service
               0.911    0.009    0.941      0.911    0.926      0.914    0.993     0.977     Stakeholder Enterprise Portal
               0.632    0.004    0.237      0.632    0.345      0.385    0.993     0.338     Burial and Memorial
Weighted Avg.  0.850    0.024    0.851      0.850    0.850      0.827    0.982     0.915

=== Confusion Matrix ===

    a     b     c     d     e     f     g     h     i   <-- classified as
 9603   117   104   170   239    29  1084    33     4 |     a = Outreach Transation and Economic Development
   72  9413   434    83   245    18   407   100    67 |     b = Home and Auto
    9   459 10210    34   131     5   205     8    34 |     c = Life Insurance
  157    61   125 12912   143    26   879   111    37 |     d = Veteran Readiness and Employment
  244   255   114   145  9456    20   614    64    59 |     e = Fiduciary Program
   54    51     2    29    59    60   106    12    20 |     f = eBenefits
 1193   309   102   972   631    51  7115   329   107 |     g = Education Service
   20   152   114   220   114    11   385 10485     3 |     h = Stakeholder Enterprise Portal
    0     6     1    13     5     0    35     0   103 |     i = Burial and Memorial
```

GainRatioAttributeEval with J48:

```
=== Summary ===

Correctly Classified Instances       60166               73.7256 %
Incorrectly Classified Instances     21442               26.2744 %
Kappa statistic                          0.6935
Mean absolute error                      0.0625
Root mean squared error                  0.2299
Relative absolute error                 32.8003 %
Root relative squared error             74.4839 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.745    0.044    0.732      0.745    0.739      0.696    0.852     0.622     Outreach Transation and Economic Development
                 0.754    0.043    0.728      0.754    0.741      0.700    0.859     0.612     Home and Auto
                 0.820    0.027    0.827      0.820    0.824      0.796    0.908     0.716     Life Insurance
                 0.800    0.042    0.803      0.800    0.801      0.759    0.880     0.705     Veteran Readiness and Employment
                 0.744    0.041    0.739      0.744    0.742      0.701    0.854     0.613     Fiduciary Program
                 0.097    0.003    0.143      0.097    0.116      0.114    0.573     0.031     eBenefits
                 0.488    0.077    0.490      0.488    0.489      0.411    0.690     0.348     Education Service
                 0.814    0.027    0.833      0.814    0.824      0.795    0.893     0.708     Stakeholder Enterprise Portal
                 0.123    0.001    0.202      0.123    0.153      0.156    0.590     0.035     Burial and Memorial
Weighted Avg.    0.737    0.043    0.736      0.737    0.737      0.694    0.848     0.619

=== Confusion Matrix ===

    a     b     c     d     e     f     g     h     i   <-- classified as
 8480   265   102   427   381    46  1547   134     1 |     a = Outreach Transation and Economic Development
  245  8170   711   243   453    43   651   309    14 |     b = Home and Auto
  126   826  9096   167   358     9   390   119     4 |     c = Life Insurance
  441   279   239 11554   369    28  1216   312    13 |     d = Veteran Readiness and Employment
  423   489   308   380  8167    26   831   335    12 |     e = Fiduciary Program
   66    57     8    43    57    38    98    21     5 |     f = eBenefits
 1625   737   356  1222   870    54  5273   645    27 |     g = Education Service
  160   387   161   329   377    20   699  9368     3 |     h = Stakeholder Enterprise Portal
   13    15    13    25    19     1    53     4    20 |     i = Burial and Memorial
```

Correlation with MLP:

```
=== Summary ===

Correctly Classified Instances       68405               83.8214 %
Incorrectly Classified Instances     13203               16.1786 %
Kappa statistic                          0.8112
Mean absolute error                      0.058
Root mean squared error                  0.167
Relative absolute error                 30.4464 %
Root relative squared error             54.0892 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.852    0.022    0.861      0.852    0.857      0.834    0.984     0.914     Outreach Transation and Economic Development
                 0.881    0.031    0.813      0.881    0.846      0.822    0.977     0.909     Home and Auto
                 0.899    0.019    0.882      0.899    0.891      0.873    0.989     0.938     Life Insurance
                 0.874    0.019    0.909      0.874    0.891      0.869    0.982     0.947     Veteran Readiness and Employment
                 0.862    0.030    0.817      0.862    0.839      0.813    0.972     0.896     Fiduciary Program
                 0.000    0.000    ?          0.000    ?          ?        0.726     0.018     eBenefits
                 0.639    0.045    0.685      0.639    0.661      0.612    0.908     0.691     Education Service
                 0.886    0.022    0.868      0.886    0.877      0.857    0.986     0.928     Stakeholder Enterprise Portal
                 0.000    0.000    ?          0.000    ?          ?        0.830     0.010     Burial and Memorial
Weighted Avg.    0.838    0.026    ?          0.838    ?          ?        0.971     0.886

=== Confusion Matrix ===

    a     b     c     d     e     f     g     h     i   <-- classified as
 9701   202    65   144   315     0   950     6     0 |     a = Outreach Transation and Economic Development
   86  9550   610    39   216     0   177   161     0 |     b = Home and Auto
   17   697  9974    29   160     0   188    30     0 |     c = Life Insurance
  142   122   116 12629   273     0   912   257     0 |     d = Veteran Readiness and Employment
  264   301   138   164  9458     0   354   292     0 |     e = Fiduciary Program
   36    56     2    21    71     0   166    41     0 |     f = eBenefits
 1014   448   305   629   769     0  6903   741     0 |     g = Education Service
    1   359    80   208   316     0   350 10190     0 |     h = Stakeholder Enterprise Portal
    1    15    14    27     4     0    83    19     0 |     i = Burial and Memorial
```

Correlation with PART:

```
=== Summary ===

Correctly Classified Instances       61008               74.7574 %
Incorrectly Classified Instances     20600               25.2426 %
Kappa statistic                          0.7052
Mean absolute error                      0.0832
Root mean squared error                  0.2081
Relative absolute error                 43.6735 %
Root relative squared error             67.4167 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.796    0.029    0.816      0.796    0.806      0.775    0.940     0.822     Outreach Transation and Economic Development
                 0.746    0.045    0.719      0.746    0.732      0.691    0.927     0.742     Home and Auto
                 0.789    0.036    0.776      0.789    0.782      0.748    0.951     0.810     Life Insurance
                 0.826    0.043    0.804      0.826    0.815      0.775    0.943     0.823     Veteran Readiness and Employment
                 0.739    0.044    0.721      0.739    0.730      0.687    0.916     0.713     Fiduciary Program
                 0.000    0.000    ?          0.000    ?          ?        0.765     0.020     eBenefits
                 0.508    0.060    0.562      0.508    0.534      0.467    0.837     0.510     Education Service
                 0.830    0.036    0.790      0.830    0.810      0.778    0.950     0.811     Stakeholder Enterprise Portal
                 0.000    0.000    ?          0.000    ?          ?        0.819     0.020     Burial and Memorial
Weighted Avg.    0.748    0.042    ?          0.748    ?          ?        0.924     0.747

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
  9066   248   189   411   384     0   926   159     0 |     a = Outreach Transation and Economic Development
   135  8091   895   238   484     0   552   444     0 |     b = Home and Auto
    74  1126  8758   187   419     0   349   182     0 |     c = Life Insurance
   285   195   254 11942   326     0  1052   397     0 |     d = Veteran Readiness and Employment
   311   458   425   380  8111     0   757   529     0 |     e = Fiduciary Program
    30    80    10    42    53     0   119    59     0 |     f = eBenefits
  1127   702   503  1310   917     0  5491   759     0 |     g = Education Service
    68   331   246   293   543     0   474  9549     0 |     h = Stakeholder Enterprise Portal
     8    22    13    52    15     0    49     4     0 |     i = Burial and Memorial
```

Correlation with Naive Bayes:

```
=== Summary ===

Correctly Classified Instances       69459               85.113  %
Incorrectly Classified Instances     12149               14.887  %
Kappa statistic                          0.8264
Mean absolute error                      0.0438
Root mean squared error                  0.1545
Relative absolute error                 22.9635 %
Root relative squared error             50.0461 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.882    0.017    0.896      0.882    0.889      0.871    0.990     0.958     Outreach Transation and Economic Development
                 0.853    0.021    0.864      0.853    0.859      0.837    0.988     0.935     Home and Auto
                 0.910    0.016    0.898      0.910    0.904      0.889    0.993     0.958     Life Insurance
                 0.892    0.025    0.885      0.892    0.889      0.865    0.987     0.958     Veteran Readiness and Employment
                 0.846    0.027    0.831      0.846    0.838      0.813    0.984     0.918     Fiduciary Program
                 0.181    0.001    0.372      0.181    0.243      0.257    0.959     0.206     eBenefits
                 0.684    0.052    0.666      0.684    0.675      0.624    0.945     0.748     Education Service
                 0.902    0.012    0.926      0.902    0.914      0.900    0.993     0.972     Stakeholder Enterprise Portal
                 0.448    0.002    0.302      0.448    0.360      0.366    0.990     0.328     Burial and Memorial
Weighted Avg.    0.851    0.024    0.851      0.851    0.851      0.827    0.983     0.919

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 10040    85   102   101   266     9   762    18     0 |     a = Outreach Transation and Economic Development
    56  9251   565    84   293    22   414   108    46 |     b = Home and Auto
    13   534 10094    33   124     1   255    23    18 |     c = Life Insurance
   105    56   116 12890   173    15   937   140    19 |     d = Veteran Readiness and Employment
   212   291   111   164  9278    18   741   140    16 |     e = Fiduciary Program
    24    50     2    31    51    71   143    17     4 |     f = eBenefits
   737   291   136  1000   765    45  7390   381    64 |     g = Education Service
    16   147   113   236   211    10   397 10372     2 |     h = Stakeholder Enterprise Portal
     0     7     1    20     2     0    60     0    73 |     i = Burial and Memorial
```

Correlation with J48:

```
=== Summary ===

Correctly Classified Instances       60269               73.8518 %
Incorrectly Classified Instances     21339               26.1482 %
Kappa statistic                          0.695
Mean absolute error                      0.0619
Root mean squared error                  0.2295
Relative absolute error                 32.4828 %
Root relative squared error             74.344  %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.802    0.036    0.781      0.802    0.792      0.757    0.890     0.696     Outreach Transation and Economic Development
                 0.738    0.041    0.733      0.738    0.735      0.695    0.857     0.603     Home and Auto
                 0.810    0.031    0.803      0.810    0.806      0.776    0.905     0.700     Life Insurance
                 0.801    0.044    0.796      0.801    0.799      0.755    0.885     0.688     Veteran Readiness and Employment
                 0.717    0.045    0.714      0.717    0.716      0.672    0.838     0.576     Fiduciary Program
                 0.089    0.003    0.112      0.089    0.099      0.096    0.585     0.024     eBenefits
                 0.502    0.073    0.513      0.502    0.507      0.433    0.697     0.367     Education Service
                 0.803    0.029    0.818      0.803    0.810      0.780    0.894     0.690     Stakeholder Enterprise Portal
                 0.117    0.001    0.216      0.117    0.151      0.157    0.621     0.035     Burial and Memorial
Weighted Avg.    0.739    0.042    0.737      0.739    0.738      0.695    0.853     0.618

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
  9131   193   133   340   406    27  1058    94     1 |     a = Outreach Transation and Economic Development
   244  7995   867   254   514    39   633   278    15 |     b = Home and Auto
   170   899  8983   163   307     5   415   149     4 |     c = Life Insurance
   368   234   210 11580   391    43  1273   337    15 |     d = Veteran Readiness and Employment
   437   519   317   411  7871    42   924   444     6 |     e = Fiduciary Program
    35    52    11    55    63    35   113    27     2 |     f = eBenefits
  1165   648   464  1351   943    79  5421   721    17 |     g = Education Service
   127   350   192   349   511    42   690  9234     9 |     h = Stakeholder Enterprise Portal
     8    19    12    38    15     0    48     4    19 |     i = Burial and Memorial
```

Analysis:

Table of Accuracy Values for Each Attribute Selection and Classification Combination						
Attribute (Right) ---------- Classification (Below)	No Attribute Removal	Principal Components Analysis	ReliefFAttributeEval	GainRatioAttributeEval	Correlation	Averages
Multilayer Perceptron	98.53%	96.63%	86.39%	83.21%	83.82%	89.72%
PART	84.42%	94.42%	79.76%	77.44%	74.76%	82.16%
Naives Bayes	96.81%	96.79%	87.22%	84.99%	85.11%	90.18%
J48	78.52%	94.35%	75.44%	73.73%	73.85%	79.18%
Averages:	89.57%	95.55%	82.20%	79.84%	79.39%	

After running 20 different classification models with 5 different Attribute Selection Algorithms combined with 4 different classification algorithms.

The best combinations were:

1. No Attribute Removal + Multilayer Perceptron (98.53%)
2. No Attribute Removal + Naives Bayes
3. Principal Component + Naives Bayes
4. Principal Component + Multilayer Perceptron

As they all achieved above a 95% accuracy.

The worst combinations were:

1. GainRatioAttributeEval + J48 (73.73%)
2. ReliefFAttributeEval + J48
3. GainRatioAttributeEval + PART
4. No Attribute Removal + J48
5. ReliefFAttributeEval + PART

After taking averages of accuracies for different attributes selections algorithms, **Principal Component Analysis** is the best for any classification task since the average is 95.55% across the four different classification methods.

Best Confusion Matrix:

```
=== Summary ===

Correctly Classified Instances       80408               98.5296 %
Incorrectly Classified Instances      1200                1.4704 %
Kappa statistic                          0.9828
Mean absolute error                      0.0075
Root mean squared error                  0.0529
Relative absolute error                  3.9617 %
Root relative squared error             17.1303 %
Total Number of Instances            81608

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.993    0.002    0.988      0.993    0.991      0.989    0.999     0.998     Outreach Transation and Economic Development
                 0.994    0.003    0.983      0.994    0.988      0.987    1.000     0.998     Home and Auto
                 0.994    0.002    0.989      0.994    0.991      0.990    1.000     0.998     Life Insurance
                 0.993    0.003    0.986      0.993    0.990      0.988    1.000     0.998     Veteran Readiness and Employment
                 0.990    0.003    0.984      0.990    0.987      0.985    0.999     0.997     Fiduciary Program
                 0.000    0.000    0.000      0.000    0.000      -0.000   0.571     0.057     eBenefits
                 0.984    0.004    0.974      0.984    0.979      0.976    0.999     0.995     Education Service
                 0.996    0.001    0.992      0.996    0.994      0.993    1.000     0.999     Stakeholder Enterprise Portal
                 0.000    0.000    ?          0.000    ?          ?        0.577     0.023     Burial and Memorial
Weighted Avg.    0.985    0.002    ?          0.985    ?          ?        0.997     0.991

=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i   <-- classified as
 11306     9     7    15    18     0    24     4     0 |     a = Outreach Transation and Economic Development
     2 10770    26    16    11     0    14     0     0 |     b = Home and Auto
     5    41 11023     6     9     0     8     3     0 |     c = Life Insurance
    14    11    13 14353     3     0    40    17     0 |     d = Veteran Readiness and Employment
    22    10    15     5 10856     1    47    15     0 |     e = Fiduciary Program
    57    44    31    73    69     0    94    25     0 |     f = eBenefits
    28    23    11    48    35     0 10637    27     0 |     g = Education Service
     1     0     3    19     8     0    10 11463     0 |     h = Stakeholder Enterprise Portal
     6    47    19    18    24     0    48     1     0 |     i = Burial and Memorial
```

If we look at all of the metrics for this classification, we can see that a majority of them are around 0.999, except the eBenefits class. We think this is because the eBenefits attribute has the least amount of instances. This would lead to the accuracy and confusion matrix for eBenefits to be much lower than other classes.

Conclusion

Based on our results, we can conclude that **No Attribute Removal + Multilayer Perceptron** had the highest accuracy, 98.53%.

This is an interesting result as this tells us removing attributes, even the ones with less correlation or miniscule variance, still has an effect on the classification accuracy. However, running the model on the data with no attributes removed took 5 times longer, which would make sense as the dataset is much longer. If attribute number must be lowered, future projects with this dataset should utilize a Principal Component Analysis to transform data because it resulted in the highest average of classification accuracies compared to other attribute selection methods.

How to reproduce our Optimal Model (No Attribute Selection + Multilayer Perceptron):

1. Open Weka and load the “embeddings_82250_attributes_v2_top100.csv” file from our project folder.
2. Click Classify in the top menu of Weka
3. Click Choose
4. Under the ‘Bayes’ folder, choose the **MultiLayerPerceptron** option
5. Click the word “**MultiLayerPerceptron**” to open the hyperparameter options
6. Choose **False** for GUI.
7. Choose **True** for autoBuild.
8. Set **100** for batchSize.
9. Choose **False** for debug.
10. Choose **False** for decay.
11. Choose **False** for doNotCheckCapabilities.
12. Set **5** for hiddenLayers.
13. Set **0.3** for learningRate.
14. Set **0.2** for momentum.
15. Choose **True** for nominalToBinaryFilter.
16. Choose **True** for normalizeAttributes.
17. Choose **True** for normalizeNumericClass.
18. Set **2** for numDecimalPlaces.
19. Choose **True** for reset.
20. Choose **False** for resume.
21. Set **0** for seed.
22. Set **8** for trainingTime.
23. Set **0** for validationSetSize.
24. Set **20** for validationThreshold.
25. Click nom (class) in the class selection area on the left middle of the page
26. Click Run

27. Wait for results

Team Members and Tasks Performed:

Anurag Perakalapudi:

- Selected and Described Attribute Selection Algorithms
- Selected and Described Classification Algorithms
- Analyzed Results
- Wrote Conclusion
- Wrote Reproduction Procedures
- Created Presentation

Aneesh Kalla:

- Created webscraper and scraped text from VA website
- Preprocessed text
- Created embeddings
- Ran WEKA models
- Wrote introduction materials

Appendix and Sources

Veterans Affairs Webpage:
Files attached with the report:

References

Weka class description (GainRatioAttributeEval)
www.knowva.ebenefits.va.gov