Principal Component Analysis and Factor Analysis for Feature Selection in Credit Rating

Shenghuan Yang  lonut Florescu
Jiangxi University of Finance and Economics  Stevens Institute of Technology

Abstract

The credit rating is an evaluation of the credit risk of a company that values the ability to pay back the debt and predict the likelihood of the debtor defaulting. There are various features influencing credit rating. Therefore it is very important to select substantive features to explore the main reason for credit rating change. To address this issue, this paper exploits Principal Component Analysis and Factor Analysis as feature selection algorithms to select important features, summarise the similar features together and obtain a minimum set of features for four sectors, Financial Sector, Energy Sector, Health Care Sector, Consumer Discretionary Sector. This paper uses two data sets, Financial Ratio and Balance Sheet, with two mappings, Detailed Mapping and Coarse Mapping, converting the target variable(credit rating) into categorical variable. To test the accuracy of credit rating prediction, Random Forest Classifier is used to test and train feature sets. The results show that the accuracy of Financial Ratio feature sets are higher than that of Balance Sheet feature sets. What is more, Factor Analysis can reduce significantly the number of features to obtain almost the same accuracy that can decrease dramatically the time spent on analyzing data. In addition, we summarise seven dominant factors and ten dominant factors having effect on credit rating change in Financial Ratio and Balance Sheet respectively by utilizing Factor Analysis which can explain the reason of credit rating change better.
1 Data

This paper uses two data sets as feature sets, Financial Ratio and Balance Sheet. Data was obtained from Wharton Research Data Services (WRDS) Database for all the S&P companies. Every sector has its own operating ways and characteristics in a company and has different effect on the credit rating of a company. Therefore, we chose specifically four typical sectors, Financial Sector, Energy Sector, Health Care Sector, Consumer Discretionary Sector. Because the type of credit rating was character variable, we needed to convert it into categorical variable with two mappings, Detailed Mapping and Coarse Mapping. The mappings were shown in Table 1.1 Table 1.2. Finally, Financial Ratio and Balance Sheet combined with two mappings to form four four data sets in every sector. The results were shown in Table 1.3.

Table 1.1: Detailed Mapping

| Rating | Mapping | Rating | Mapping |
|--------|---------|--------|---------|
| AAA    | 2       | BB-    | 13      |
| AA+    | 3       | B+     | 14      |
| AA     | 4       | B      | 15      |
| AA-    | 5       | B-     | 16      |
| A+     | 6       | CCC+   | 17      |
| A      | 7       | CCC    | 18      |
| A-     | 8       | CCC    | 19      |
| BBB+   | 9       | CC     | 20      |
| BBB    | 10      | C      | 21      |
| BBB-   | 11      | RD     | 22      |

Table 1.2: Coarse Mapping

| Rating         | Mapping | Rating Description     |
|----------------|---------|------------------------|
| AAA            | 1       | Prime                  |
| AA+, AA, AA-   | 2       | High Grade             |
| A+, A, A-      | 3       | Upper Medium Grade     |
| BBB+, BBB, BBB-| 4       | Lower Medium Grade     |
| BB+, BB, BB-   | 5       | Non-Investment Grade Speculative |
| B+, B, B-     | 6       | Highly Speculative     |
| CCC+, CCC, CCC-| 7       | Substantial Risks      |
| CC            | 8       | Extremely Speculative  |
| C             | 9       | Default Inminent       |
| RD, SD, D     | 10      | In Default             |
2 Algorithm

2.1 Algorithm 1

Principal Component Analysis (PCA) using the idea of dimension reduction (linear transformation), a set of variables of possibly related are converted into a set of values of linearly uncorrelated variables (principal component) under the premise of less loss of information. The first principal component has the largest possible variance (taking into account the variability of the data), and each subsequent component has the highest possible variance under the constraint of orthogonal with the previous component. The resulting vectors (each of which is a linear combination of variables and contain n variables) are unrelated orthogonal basis sets. Each principal component is a linear combination of the original variables, and the principal components are not related to each other. In general, people select principal components with cumulative proportion of variance more than 85% because most of information is included.

The top reasons for Principal component analysis are:

- There are same numbers of variables and components in PCA. It can reduce the numbers of components if we just extract the components with variance cumulative proportion more than 85% that makes the follow-up analysis easier.

- The objective economic phenomenon is scientifically evaluated by calculating the score of the comprehensive principal component function.

- It focuses on the comprehensive evaluation of the influence of information contribution.

\[
\begin{align*}
X_1 &= a_{11}F_1 + a_{12}F_2 + \cdots + a_{1n}F_n \\
X_2 &= a_{21}F_1 + a_{22}F_2 + \cdots + a_{2n}F_n \\
\vdots \\
X_n &= a_{n1}F_1 + a_{n2}F_2 + \cdots + a_{nn}F_n
\end{align*}
\]  

(1)
$X_n$ is the $n_{th}$ features. $F_n$ is the $n_{th}$ components. $a_{nn} = \sqrt{\lambda_n} e_{nn}$ is the loadings of components. $\lambda_n$ is the variance of $n_{th}$ component(also eigenvalue). $\|e_n\| = 1$ is the eigenvector of $n_{th}$ component.

Before performing the PCA, we must standardize data to avoid the impact of the unit different. And in this standardized data set, the eigenvectors are unit eigenvectors. Therefore, we can regard $e_{ni}^2$ as the proportion of variance of $i_{th}$ feature of $n_{th}$ component. We select components with cumulative proportion of variance more than 85%.

\[
\text{count: the number of components when cumulative proportion of variance is more than 85 \%)}
\]

\[
X_1 = a_{11} F_1 + a_{12} F_2 + \cdots + a_{1\text{count}} F_{\text{count}} \\
X_2 = a_{21} F_1 + a_{22} F_2 + \cdots + a_{2\text{count}} F_{\text{count}} \\
\ldots \\
X_n = a_{n1} F_1 + a_{n2} F_2 + \cdots + a_{n\text{count}} F_{\text{count}}
\]

(2)

The loadings of PCA represent the relationship between components and features. The higher the loading, the more the relationship between the two variables. Therefore, we consider adding the loadings of every component of each feature to get the loading scores that represents the extent to which features affect. Then the feature set is sorted in the decreasing order of components scores. In every iteration, the features from this set are added one in order at a time, while maintaining the order, to form the data set on which the Random Forest Classifier is trained and tested on and the accuracy of each iteration is recorded.

There are positive and negative loadings that means the loadings have positive or negative effect on components. The added scores would be counteracted if the features had positive and negative loadings that we cannot test the importance of features. To avoid this case, we propose two algorithms:

\[
\text{Algorithm 1 : } Score_i = \sum_{j=1}^{\text{count}} |a_{ij}| = |\sqrt{\lambda_1} e_{i1}| + |\sqrt{\lambda_2} e_{i2}| + \cdots + |\sqrt{\lambda_{\text{count}}} e_{i\text{count}}| 
\]

\[
\text{Algorithm 2 : } Score_i = \sum_{j=1}^{\text{count}} a_{ij}^2 = \lambda_{1} e_{i1}^2 + \lambda_{2} e_{i2}^2 + \cdots + \lambda_{\text{count}} e_{i\text{count}}^2
\]

Algorithm 1 is to add the absolute value of loading of every component of each features, the Algorithm 2 is to add the square of loading. They all can avoid positive and negative offsetting and represent that features have effect on components. In order to figure out which algorithm is better, we test the accuracy of the credit rating change of Financial Data and Balance Sheet.
Figure 2.1: Algorithm Comparison in Detailed Mapping

(a) Financial Sector+Detailed Mapping  
(b) Energy Sector+Detailed Mapping  
(c) Health Sector+Detailed Mapping  
(d) Consumer Sector+Detailed Mapping

Figure 2.2: Algorithm Comparison in Coarse Mapping

(a) Financial Sector+Coarse Mapping  
(b) Energy Sector+Coarse Mapping  
(c) Health Sector+Coarse Mapping  
(d) Consumer Sector+Coarse Mapping
From Figure 2.1, 2.2, the blue and gray line show the accuracy of algorithm 1 in Financial Ratio and Balance Sheet respectively; orange and yellow line display the accuracy of algorithm 2 in Financial Ratio and Balance Sheet respectively. In Financial Ratio, there is slight difference between these two algorithms in four sectors. In Balance Sheet, yellow line (algorithm 2) is always above the gray line (algorithm 1) that means the accuracy of algorithm 2 is more accurate than that of algorithm 1 in four sectors.

However, the numbers of selected features for four data sets in every sector are different from each other because we select the features with Chi-square p-value less than 0.05 as final feature sets for four data sets in every sector. So we try to test the accuracy with the same numbers. From Figure 2.1 2.2, most accuracy curves tent to steady when the number is 20 or more. We calculate the average accuracy of the top 20 features by utilizing Random Forest that trained and tested the accuracy of feature sets for 100 times. The results are shown in Table 2.4, Table 2.5, Table 2.6, Table 2.7.

Table 2.4: Accuracy for Financial Sector

|                | Financial Ratio Detailed Mapping | Balance Sheet Detailed Mapping | Financial Ratio Coarse Mapping | Balance Sheet Coarse Mapping |
|----------------|---------------------------------|-------------------------------|--------------------------------|-------------------------------|
| Avg.Test Accuracy | 0.97169                          | 0.87244                       | 0.98999                        | 0.8964                        |
| Accuracy with Algorithm 1 using The Top 20 Features | 0.97577                          | 0.92332                       | 0.98989                        | 0.9274                        |

Table 2.5: Accuracy for Energy Sector

|                | Financial Ratio Detailed Mapping | Balance Sheet Detailed Mapping | Financial Ratio Coarse Mapping | Balance Sheet Coarse Mapping |
|----------------|---------------------------------|-------------------------------|--------------------------------|-------------------------------|
| Avg.Test Accuracy | 0.97271                          | 0.90403                       | 0.98915                        | 0.98205                       |
| Accuracy with Algorithm 1 using The Top 20 Features | 0.97035                          | 0.91974                       | 0.98947                        | 0.98064                       |
Table 2.6: Accuracy for Health Sector

|                      | Financial Ratio Detailed Mapping | Balance Sheet Detailed Mapping | Financial Ratio Coarse Mapping | Balance Sheet Coarse Mapping |
|----------------------|----------------------------------|-------------------------------|-------------------------------|------------------------------|
| Avg. Test Accuracy   | 0.97083                          | 0.86841                       | 0.99001                       | 0.95187                      |
| Accuracy with        |                                  |                               |                               |                              |
| Algorithm 1 using    |                                  |                               |                               |                              |
| The Top 20 Features  |                                  |                               |                               |                              |
| Avg. Test Accuracy   | 0.97288                          | 0.93562                       | 0.99035                       | 0.93103                      |
| Accuracy with        |                                  |                               |                               |                              |
| Algorithm 2 using    |                                  |                               |                               |                              |
| The Top 20 Features  |                                  |                               |                               |                              |

Table 2.7: Accuracy for Consumer Sector

|                      | Financial Ratio Detailed Mapping | Balance Sheet Detailed Mapping | Financial Ratio Coarse Mapping | Balance Sheet Coarse Mapping |
|----------------------|----------------------------------|-------------------------------|-------------------------------|------------------------------|
| Avg. Test Accuracy   | 0.96281                          | 0.80097                       | 0.98614                       | 0.86128                      |
| Accuracy with        |                                  |                               |                               |                              |
| Algorithm 1 using    |                                  |                               |                               |                              |
| The Top 20 Features  |                                  |                               |                               |                              |
| Avg. Test Accuracy   | 0.96734                          | 0.93562                       | 0.99035                       | 0.93103                      |
| Accuracy with        |                                  |                               |                               |                              |
| Algorithm 2 using    |                                  |                               |                               |                              |
| The Top 20 Features  |                                  |                               |                               |                              |

From Table 2.4, Table 2.5, Table 2.6, Table 2.7, most accuracy of Algorithm 2 is higher than that of algorithm 1 in four sectors. Therefore, algorithm 2 in this part is designed for the final Algorithm 1. After obtaining the final algorithm 1, a summary of this algorithm is shown in Figure 2.3.
Figure 2.3: Flowchart for Algorithm 1

1. Facilitate the Chi-square Test between each feature and target variable.
2. Record the component loadings, variance and cumulative proportion.
3. Select the principal components with the variance cumulative contribution rate of 85% or more.
4. Add the square of loadings of every component of each feature to obtain component scores.
5. Iterate over the sorted set of features and add one feature in every iteration from this set in random order.
6. Record the average accuracy $n_{th}$ features in random order.
7. Record the accuracy in every iteration.
8. Sort the features in decreasing orders of component scores.
9. Perform the principal component model.
10. Standardize data to reduce the impact of unit difference of features.
11. Select the features with the p-value less than 0.05 as final feature sets.
12. Only choose the features that lead to an increase in the accuracy.
2.2 Algorithm 2

Factor analysis is to identify the underlying relationships between measured variables. And Factor analysis is related to principal component analysis (PCA), but the two are not identical. Factor Analysis is based on the idea of dimension reduction and the study of the dependence within the correlation matrix of the original variables, some variables with complex relationships are expressed as a linear combination of a small number of common factors and special factors that only work on one variable. In this algorithm, we try to summarize the similar features together to extract common factors that can interpret features better.

The top reasons for Factor analysis are:

• Factor analysis make the factor better explained by exploiting factor rotation. The factor analysis is more dominant in the interpretation of the principal component.

• Factor analysis is not to delete original variables, but the information of the original variable is recombined to find a common factor affecting the target variable and simplify data.

Before performing Factor Analysis, feature set must be tested by KMO and Bartlett’s Test to examine whether the data is suitable for factor analysis.

The Kaiser-Meyer-Olkin (KMO) test measures the applicability of our data to factor analysis. The test measures sampling adequacy for each variable in the model and for the complete model. KMO values less than 0.6 indicate the sampling is not adequate that factor analysis should not be taken.

Bartlett’s Test of Sphericity is based on the correlation coefficient matrix. Its null hypothesis correlation coefficient matrix is a unit array, that is, all the elements of the diagonal of the correlation coefficient matrix are 1, and all the non-diagonal elements are zero. The statistics of Bartlett’s Test of Sphericity is derived from the determinant of the correlation coefficient matrix. If the value is large, and the corresponding concomitant probability value is less than the specified significant level, the null hypothesis is rejected, indicating that the correlation coefficient matrix is not a unit array, there is a correlation between the original variables. Small p-value (less than 0.05) of the significance level indicate that a factor analysis may be useful with data.

Figure 2.4: KMO and Bartlett’s Test

| Financial Ratio | Balance Sheet |
|-----------------|---------------|
| **KMO**        | Financial     | Energy | Health | Consumer |
| Bartlett(p-value) | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| **KMO**        | 0.72    | 0.8    | 0.77   | 0.66   |
| Bartlett(p-value) | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| **KMO**        | 0.72    | 0.5    | 0.83   | 0.63   |

(a) Financial Ratio  
(b) Balance Sheet
From Figure 2.4, almost all the p-values of Bartlett’s Test are less than 0.05 and KMO values are more than 0.6 in four sectors in Financial Ratios and Balance Sheet except for Energy Sector in Balance Sheet.

After performing the Factor Analysis to get the rotated loadings, the features with loadings more than 0.5 are selected and the features are deleted if their loadings with less than 0.5. The feature represents the factor that its loading is maximum if they have two or more loadings more than 0.5. Because the factors are unrelated to each other, each feature can be only included by one factor, there are no same features between factors. The loading of features of factor means that have positive or negative effect on factor and there is first factor, second factor and so on with their effects gradually weakening. According to the principle of priority of the first factor, the features of each factor are sorted according to the absolute value of loadings within first factor and then second factor and so on.

For example, if first factor includes three features, their order are 1, 2, 3; then second factors has four features, their order are 4, 5, 6, 7 and so on. To obtain this sorted set, the algorithm 2 is proposed. If there are m factors, feature A is included by the $j_{th}$ factor, the absolute value of loading of A plus (m-$j$+1) to get factor scores. It means that absolute value of loadings of the features of first factor plus highest score m, that of the second factor plus second highest score (m-1), that of the $j_{th}$ factor plus score (m-$j$+1) and so on. Then the feature set is sorted in the decreasing order of added scores. In every iteration, the features from this set are added one in order at a time, while maintaining the order, to form the data set on which the Random Forest Classifier is trained and tested on and the accuracy of each iteration is recorded. A summary of this algorithm is shown in Figure 2.5.
Figure 2.5: Flowchart for Algorithm 2

1. Facilitate the Chi-square Test between each feature and target variable.
2. Record the p-values of all features.
3. Selected the features with the p-value less than 0.05 as final feature sets.
4. Selected features that their loadings of factors more than 0.5.
5. Performed Factor Analysis model and obtained rotated factor loadings.
6. Performed KMO and Bartlett’s Test to examine whether the feature sets are suitable for factor analysis.
7. Selected the highest loadings if features have two or more loadings more than 0.5.
8. Sorted the features that are included by the first factor, second factor, third factor and so on.
9. Classified the features of each factor.
10. Selected the top 10 features.
11. Sort the features in decreasing orders of added loadings.
12. The absolute value of loading of features of each factor add \((m-j+1)(m: \text{number of factors}, j: \text{the } j^{th} \text{factor})\).
13. Iterate over the sorted set of features and add one feature in every iteration from this set in order.
14. Iterate over the sorted set of features and add one feature in every iteration from this set in random order.
15. Only choose the features that lead to an increase in the accuracy.
16. Compared with the accuracy of sorted set and random set to examine whether the algorithm is better for predicting credit rating.
17. Calculate the average accuracy in random order.
18. Record the accuracy in every iteration.
3 Result

3.1 Accuracy Results

In Algorithm 1, the components with cumulative variance proportion more than 85% were selected as final principal components. The number of component was shown in Table 3.1. In Algorithm 2, the factors with cumulative variance proportion more than 85% were selected as final factors. The number of factor was shown in Table 3.2.

In this section, the results for both the algorithms will be summarized according to sectors. We compared the accuracy of the top 20 features selected by Algorithm 1 and the accuracy of the features selected by Algorithm 2. These two feature sets were trained and tested for 100 times by utilizing Random Forest Classifier and the average accuracy was recorded. The results were shown in Table 3.3 Table 3.4 Table 3.5 Table 3.6

Table 3.1: The Number of Component for Algorithm 1

| Financial Ratio | Financial Ratio | Balance Sheet | Balance Sheet |
|-----------------|-----------------|---------------|---------------|
| Detailed Mapping| Coarse Mapping  | Detailed Mapping| Coarse Mapping |
| Financial Sector| 22              | 22            | 33            | 32            |
| Energy Sector   | 18              | 18            | 22            | 21            |
| Health Sector   | 22              | 22            | 29            | 32            |
| Consumer Sector | 28              | 29            | 28            | 28            |

Table 3.2: The Number of Factors for Algorithm 2

| Financial Ratio | Financial Ratio | Balance Sheet | Balance Sheet |
|-----------------|-----------------|---------------|---------------|
| Detailed Mapping| Coarse Mapping  | Detailed Mapping| Coarse Mapping |
| Financial Sector| 10              | 10            | 15            | 14            |
| Energy Sector   | 8               | 8             | NA            | NA            |
| Health Sector   | 11              | 10            | 11            | 13            |
| Consumer Sector | 11              | 11            | 13            | 12            |
Table 3.3: Accuracy Results for Financial Sector

|                  | Detailed Mapping | Detailed Mapping | Coarse Mapping | Coarse Mapping |
|------------------|------------------|------------------|---------------|---------------|
| Avg. Test Accuracy |                  |                  |               |               |
| Accuracy with Algorithm 1 using The Top 20 Features | 0.97577 | 0.92332 | 0.98989 | 0.9274 |
| Avg. Test Accuracy with Algorithm 2 | 0.97465 | 0.92331 | 0.98983 | 0.95227 |
| Avg. No feature Selection | 0.97712 | 0.92766 | 0.99203 | 0.96721 |

Table 3.4: Accuracy Results for Energy Sector

|                  | Detailed Mapping | Detailed Mapping | Coarse Mapping | Coarse Mapping |
|------------------|------------------|------------------|---------------|---------------|
| Avg. Test Accuracy |                  |                  |               |               |
| Accuracy with Algorithm 1 using The Top 20 Features | 0.97035 | 0.91974 | 0.98947 | 0.98064 |
| Avg. Test Accuracy with Algorithm 2 | 0.96859 | NA | 0.98831 | NA |
| Avg. No feature Selection | 0.97253 | 0.93115 | 0.99027 | 0.98647 |

Table 3.5: Accuracy Results for Health Sector

|                  | Detailed Mapping | Detailed Mapping | Coarse Mapping | Coarse Mapping |
|------------------|------------------|------------------|---------------|---------------|
| Avg. Test Accuracy |                  |                  |               |               |
| Accuracy with Algorithm 1 using The Top 20 Features | 0.97288 | 0.93562 | 0.99035 | 0.93103 |
| Avg. Test Accuracy with Algorithm 2 | 0.97363 | 0.93091 | 0.99081 | 0.96093 |
| Avg. No feature Selection | 0.97402 | 0.94367 | 0.99112 | 0.96389 |
3.2 Important Features Selected by Algorithm 1

To compare two algorithms, the accuracy of top 20 features selected by algorithm 1 was calculated for every data set in every sector. However, the features numbers are more than 20, all features of every sector for each data set were listed according to the component scores by a bar plot in this section.

3.2.1 Financial Sector

This section displays the important features by Algorithm 1 for the Financial Sector for four data sets, Financial Ratio+Detailed Mapping, Balance Sheet+Detailed Mapping, Ratio+Coarse Mapping, Balance Sheet+Coarse Mapping.

Figure 3.1: Financial Sector: Important Features in Financial Ratio+Detailed Mapping
Figure 3.2: Financial Sector: Important Features in Balance Sheet+Detailed Mapping

Figure 3.3: Financial Sector: Important Features in Financial Ratio+Coarse Mapping
3.2.2 Energy Sector

This section displays the important features by Algorithm 1 for the Energy Sector for four data sets, Financial Ratio+Detailed Mapping, Balance Sheet+Detailed Mapping, Ratio+Coarse Mapping, Balance Sheet+Coarse Mapping.
Figure 3.5: Energy Sector: Important Features in Financial Ratio+Detailed Mapping

Figure 3.6: Energy Sector: Important Features in Balance Sheet+Detailed Mapping
Figure 3.7: Energy Sector: Important Features in Financial Ratio+Coarse Mapping

Figure 3.8: Energy Sector: Important Features in Balance Sheet+Coarse Mapping
3.2.3 Health Sector

This section displays the important features by Algorithm 1 for the Health Sector for four data sets, Financial Ratio+Detailed Mapping, Balance Sheet+Detailed Mapping, Ratio+Coarse Mapping, Balance Sheet+Coarse Mapping.

Figure 3.9: Health Sector: Important Features in Financial Ratio+Detailed Mapping
Figure 3.10: Health Sector: Important Features in Balance Sheet+Detailed Mapping

Figure 3.11: Health Sector: Important Features in Financial Ratio+Coarse Mapping
3.2.4 Consumer Sector

This section displays the important features by Algorithm 1 for the Consumer Sector for four data sets, Financial Ratio+Detailed Mapping, Balance Sheet+Detailed Mapping, Ratio+Coarse Mapping, Balance Sheet+Coarse Mapping.
Figure 3.13: Consumer Sector: Important Features in Financial Ratio+Detailed Mapping

Figure 3.14: Consumer Sector: Important Features in Balance Sheet+Detailed Mapping
Figure 3.15: Consumer Sector: Important Features in Financial Ratio+Coarse Mapping

Figure 3.16: Consumer Sector: Important Features in Balance Sheet+Coarse Mapping
3.3 Important Feature Selected by Algorithm 2

3.3.1 Financial Sector

This section displays the important features by Algorithm 2 for the Consumer Sector for four data sets, Financial Ratio+Detailed Mapping, Balance Sheet+Detailed Mapping, Ratio+Coarse Mapping, Balance Sheet+Coarse Mapping.

Figure 3.17: Financial Sector: Important Features in Financial Ratio+Detailed Mapping
Figure 3.18: Financial Sector: Important Features in Balance Sheet+Detailed Mapping

Figure 3.19: Financial Sector: Important Features in Financial Ratio+Coarse Mapping
3.3.2 Energy Sector

This section displays the important features by Algorithm 2 for the Energy Sector for four data sets, Financial Ratio+Detailed Mapping, Balance Sheet+Detailed Mapping, Ratio+Coarse Mapping, Balance Sheet+Coarse Mapping.
3.3.3 Health Sector

This section displays the important features by Algorithm 2 for the Health Sec-
tor for four data sets, Financial Ratio+Detailed Mapping, Balance Sheet+Detailed Mapping, Ratio+Coarse Mapping, Balance Sheet+Coarse Mapping.

Figure 3.23: Health Sector: Important Features in Financial Ratio+Detailed Mapping
Figure 3.24: Health Sector: Important Features in Balance Sheet+Detailed Mapping

Figure 3.25: Health Sector: Important Features in Financial Ratio+Coarse Mapping
3.3.4 Consumer Sector

This section displays the important features by Algorithm 2 for the Consumer Sector for four data sets, Financial Ratio+Detailed Mapping, Balance Sheet+Detailed Mapping, Ratio+Coarse Mapping, Balance Sheet+Coarse Mapping.
Figure 3.27: Consumer Sector: Important Features in Financial Ratio+Detailed Mapping

Figure 3.28: Consumer Sector: Important Features in Balance Sheet+Detailed Mapping
Figure 3.29: Consumer Sector: Important Features in Financial Ratio+Coarse Mapping

Figure 3.30: Consumer Sector: Important Features in Balance Sheet+Coarse Mapping
4 Discussion and Conclusion

4.1 Discussion

We can see from Table 2.5, 3.4, 3.5, 3.6, they showed the similar results in terms of the accuracy of Algorithms. However, the number of selected features by them are different. For all the original feature sets, the features with the p-value less than 0.05 were selected by exploiting Chi-square Test. Algorithm 1 was to add the loading square of every component for each features as component scores of each features and sort the features according to the component scores, it did not need to delete features. Therefore the number of feature sets of Algorithm 1 was equal to that of Chi-Square Test. Algorithm 2 was to delete features that had no loadings with more than 0.5 and sort the deleted feature sets. Therefore the number of feature sets of Algorithm 2 was not equal to that of Chi-Square Test. The results are shown in Table 4.1, 4.2, 4.3, 4.4.

Table 4.1: The Number of original Features

|                | Financial Ratio Detailed Mapping | Financial Ratio Coarse Mapping | Balance Sheet Detailed Mapping | Balance Sheet Coarse Mapping |
|----------------|----------------------------------|-------------------------------|-------------------------------|-------------------------------|
| Financial Sector | 52                               | 52                            | 101                           | 101                           |
| Energy Sector   | 57                               | 57                            | 111                           | 111                           |
| Health Sector   | 59                               | 59                            | 116                           | 116                           |
| Consumer Sector | 58                               | 58                            | 103                           | 103                           |

Table 4.2: The Number of Features selected by Chi-Square Test

|                | Financial Ratio Detailed Mapping | Financial Ratio Coarse Mapping | Balance Sheet Detailed Mapping | Balance Sheet Coarse Mapping |
|----------------|----------------------------------|-------------------------------|-------------------------------|-------------------------------|
| Financial Sector | 51                               | 50                            | 76                            | 74                           |
| Energy Sector   | 54                               | 55                            | 71                            | 68                           |
| Health Sector   | 57                               | 56                            | 75                            | 82                           |
| Consumer Sector | 58                               | 57                            | 73                            | 66                           |

Table 4.3: The Number of Features for Algorithm 1

|                | Financial Ratio Detailed Mapping | Financial Ratio Coarse Mapping | Balance Sheet Detailed Mapping | Balance Sheet Coarse Mapping |
|----------------|----------------------------------|-------------------------------|-------------------------------|-------------------------------|
| Financial Sector | 51                               | 50                            | 76                            | 74                           |
| Energy Sector   | 54                               | 55                            | 71                            | 68                           |
| Health Sector   | 57                               | 56                            | 75                            | 82                           |
| Consumer Sector | 58                               | 57                            | 73                            | 66                           |
Table 4.4: The Number of Features selected for Algorithm 2

|                     | Financial Ratio | Financial Ratio | Balance Sheet | Balance Sheet |
|---------------------|-----------------|-----------------|---------------|---------------|
|                     | Detailed Mapping| Coarse Mapping   | Detailed Mapping| Coarse Mapping|
| Financial Sector    | 31              | 31              | 52            | 50            |
| Energy Sector       | 37              | 38              | NA            | NA            |
| Health Sector       | 36              | 35              | 52            | 57            |
| Consumer Sector     | 29              | 30              | 48            | 43            |

From Table 4.1, Table 4.2, Table 4.3, Table 4.4, the number of feature sets of Algorithm 1 was equal to that of Chi-Square Test. The number of features selected by Algorithm 2 was less than that of Chi-square Test. The features sets were sorted in decreasing order of components scores/factors scores and were carried out the iteration adding one feature in order in every time. We can obtain the minimum number of features when the accuracy curves became steady. To examine whether the sorted features better, the random order feature of every data set also did this iteration and repeated 10 times (each time a feature set formed a new random feature set) and the average accuracy of $n_{th}$ feature is recorded. The accuracy of comparison plot about sorted feature sets and random feature sets was displayed in Figure 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8.

Table 4.5: The Number of Features Shortlisted by Algorithm 1

|                     | Financial Ratio | Financial Ratio | Balance Sheet | Balance Sheet |
|---------------------|-----------------|-----------------|---------------|---------------|
|                     | Detailed Mapping| Coarse Mapping   | Detailed Mapping| Coarse Mapping|
| Financial Sector    | 5               | 12              | 54            | 14            |
| Energy Sector       | 10              | 10              | 19            | 19            |
| Health Sector       | 6               | 8               | 66            | 8             |
| Consumer Sector     | 33              | 8               | 50            | 40            |

Table 4.6: The Number of Features Shortlisted by Algorithm 2

|                     | Financial Ratio | Financial Ratio | Balance Sheet | Balance Sheet |
|---------------------|-----------------|-----------------|---------------|---------------|
|                     | Detailed Mapping| Coarse Mapping   | Detailed Mapping| Coarse Mapping|
| Financial Sector    | 5               | 8               | 21            | 10            |
| Energy Sector       | 8               | 13              | NA            | NA            |
| Health Sector       | 9               | 18              | 13            | 13            |
| Consumer Sector     | 7               | 13              | 32            | 37            |
Table 4.7: The Number of Random Features Shortlisted by Algorithm 1

| Sector       | Financial Ratio Detailed Mapping | Financial Ratio Coarse Mapping | Balance Sheet Detailed Mapping | Balance Sheet Coarse Mapping |
|--------------|----------------------------------|-------------------------------|-------------------------------|-----------------------------|
| Financial    | 11                               | 14                            | 66                            | 19                          |
| Energy       | 8                                | 7                             | 34                            | 34                          |
| Health       | 8                                | 12                            | 48                            | 17                          |
| Consumer     | 38                               | 10                            | 57                            | 45                          |

Table 4.8: The Number of Random Features Shortlisted by Algorithm 2

| Sector       | Financial Ratio Detailed Mapping | Financial Ratio Coarse Mapping | Balance Sheet Detailed Mapping | Balance Sheet Coarse Mapping |
|--------------|----------------------------------|-------------------------------|-------------------------------|-----------------------------|
| Financial    | 6                                | 9                             | 47                            | 39                          |
| Energy       | 8                                | 13                            | NA                            | NA                          |
| Health       | 8                                | 17                            | 49                            | 49                          |
| Consumer     | 6                                | 11                            | 36                            | 34                          |

Figure 4.1: Feature number in Financial Ratio Detailed Mapping for Algorithm 1
Figure 4.2: Feature Number in Balance Sheet Detailed Mapping for Algorithm1

(a) Financial Sector  
(b) Energy Sector  
(c) Health Sector  
(d) Consumer Sector

Figure 4.3: Feature Number in Financial Ratio Coarse Mapping for Algorithm1

(a) Financial Sector  
(b) Energy Sector  
(c) Health Sector  
(d) Consumer Sector
Figure 4.4: Feature Number in Balance Sheet Coarse Mapping for Algorithm1

(a) Financial Sector
(b) Health Sector

(c) Financial Sector
(d) Health Sector

Figure 4.5: Feature number in Financial Ratio Detailed Mapping for Algorithm2

(a) Financial Sector
(b) Energy Sector

(c) Health Sector
(d) Consumer Sector
Figure 4.6: Feature Number in Balance Sheet Detailed Mapping for Algorithm2

(a) Financial Sector  
(b) Health Sector  
(c) Consumer Sector

Figure 4.7: Feature number in Financial Ratio+Coarse Mapping for Algorithm2

(a) Financial Sector  
(b) Energy Sector  
(c) Health Sector  
(d) Consumer Sector
4.2 Conclusion

From table 4.5, table 4.6, Algorithm 2 needed lesser number of features in comparison to Algorithm 1 for most of data sets to achieve almost the same accuracy that reduced significantly the time spent on data analyzing and improved computational efficiency. In addition, the accuracy of Financial Ratio feature sets was always higher than that of Balance Sheet feature sets. From table 4.7, table 4.8, random order data sets need more number of features compared with sorted feature sets. This confirmed sorting feature sets before feature selecting was necessary that improved significantly computational efficiency.

5 Feature Classification

Factor Analysis can deal with data sets where there are a large number of observed variables that are thought to reflect a smaller number of latent variables. There are a large number of features that influence the credit rating change. Factor analysis can help to summarize similar features together to find out the latent factors that create a commonality and explain one public factor. In this Algorithm, we can reduce significantly the number of variables and obtain main factors that have effect on credit rating change to make people understand the reason of credit rating change better.
The KMO value of Energy Sector in Balance sheet was 0.5 (less than 0.6) that indicated it was not adequate for Factor Analysis. Therefore, we do not perform Factor Analysis for Energy Sector in Balance Sheet. According to four data set for four sectors, we summarized the main factors that influenced the credit rating and displayed the results by summarizing the factors of Financial Ratio and Balance Sheet for four sectors. Detailed factor results show in Appendices.

5.1 Financial Ratio

In this section, we summarized the Common factors in Financial Ratio for four sectors. There are seven factors that have influence on credit rating, Debt/Asset and Debt/capital Factor, Earning and Profit Factor, Cash Turnover Factor, Liquidity Factor, Stocks Value Factor, Financing Factor, Enterprise Value Factor.

5.1.1 Financial Ratio: Factor 1

The first factor is Debt/Asset and Debt/capital Factor. This factor includes Long-term Debt/Invested Capital, Total Debt/Capital, Total Debt/Total Assets, Total Debt/Total Assets 1, Long-term Debt/Book Equity, Long-term Debt/Total Liabilities, Total Debt/Equity.

The Debt to Asset ratio indicates the company’s financial leverage and tells people the percentage of the company’s asset financed by debt, rather than equity. The higher the ratio, the higher the degree of leverage and financial risk. The ratio can be used for investors to evaluate whether the company is reliable to meet its current debt obligation and pay a return on their investment. The debt to capital ratio gives analysts and investors a better understanding of the company’s financial structure and whether the company is worth investing. Therefore, the higher the Debt/Asset and Debt/capital, the lower the credit rating.
5.1.2 Financial Ratio: Factor 2

The second factor is Earning and Profit Factor. This factor includes Pre-tax Return on Total Earning Assets, Return on Capital Employed, After-tax Return on Average Common Equity, Return on Equity, Return on Assets, After-tax Return on Invested Capital, Operating Profit Margin After Depreciation, Net Profit Margin, Gross Profit /Total Assets, Profit /Before Depreciation Current Liabilities, Operating Profit Margin Before Depreciation.

Returns are the earnings acquired before the interest is paid. Profit mainly included the gross profit margin, operating profit margin, and net profit margin. Profit is the money that company paid for all expense in production, sales, taxes and so on. They are the indicator whether the company has enough funds to meet its current debt obligations and can run well in the long-run.
5.1.3 Financial Ratio: Factor 3

The third factor is Cash Turnover Factor. This factor includes Asset Turnover, Receivables Turnover, Inventory /Current Assets, Payables / Turnover. Turnover is an accounting concept that calculates the speed at which a company operates. Turnover ratio calculates the speed at which a company collects cash from accounts receivable and inventory investments. Analysts and investors use these ratios to determine whether the company is seen as a good investment.
5.1.4 Financial Ratio: Factor 4

This factor is Liquidity Factor. This factor includes Cash Ratio, Cash Balance/Total Liabilities, Quick Ratio (Acid Test), Current Ratio. The cash ratio is a measure of company liquidity, especially the ratio of company cash and cash equivalents to its current liabilities. The cash ratio is a liquidity indicator of the ability of companies to use only cash and cash equivalents to cover their short-term debt.

5.1.5 Financial Ratio: Factor 5

This factor is Stocks Value Factor. This factor includes P/E(Diluted,Excl,El), P/E(Diluted,Incl,El). The P / E ratio is the valuation ratio of its current share price relative to its earnings per share. Investors and analysts use the P / E ratio to determine the relative value of the company’s shares for comparison with the company.
5.1.6 Financial Ratio: Factor 6

This factor is Financing Factor. This factor includes Short-Term Debt/Total Debt, Long-term Debt/Total Liabilities, Current Liabilities/Total Liabilities. Total liabilities are the sum of individual or company debts. They are generally divided into three categories: short-term liabilities, long-term liabilities and other liabilities. Short-Term Debt/Total Debt indicates the extent of a company’s reliance on short-term financing. Long-term Debt/Total Liabilities shows the degree of a company’s reliance on long-term financing. The larger the proportion of debt, the higher the financial leverage of the company, the less stable the company has its own capital, and the need to consider paying off the debt at any time, which will require higher cash flow of the company, so that the operating risk of the company will increase.

5.1.7 Financial Ratio: Factor 7

This factor is Enterprise Value Factor. This factor includes Enterprise Value Multiple, Total Debt/EBITDA. Enterprise multiple is a ratio used to evaluate
the value of a company. Enterprise multiples are calculated by EBITDA according to the value of the enterprise. EBITDA is an abbreviation for earnings before interest, tax, depreciation and amortization. Enterprise multiples is used to determine whether a company is undervalue or overvalue for investors. The lower ratio, the company is more likely to be undervalue. The higher ratio, the company is more likely to be overvalue.

Figure 5.7: Financial Ratio: Factor 7

5.2 Balance Sheet

In this section, we summarized the Common factors in Balance Sheet for four sectors. There are ten factors influencing the credit rating, Asset and liabilities Factor, Income Factor, Implied Option Factor, Post Retirement Adjustment Factor, Pension Adjustment Factor, Earing Per Share Factor, Discontinued Operations Factor, Preference Dividends Factor.

5.2.1 Balance Sheet: Factor 1

The first factor is Asset and Liabilities Factor. This factor includes Current Assets -Other - Total, Current Liabilities -Other - Total. Current assets can reflect the operation of the company that they can be used to support daily business operation and to pay for operating expenses. The current liabilities refer to the debts or liabilities that the Company has due within one year or within the normal operating period. In addition, current liabilities are settled through the use of current assets, such as cash. Therefore, this factor represent the state of operation for a company.
5.2.2 Balance Sheet: Factor 2

The second factor is Income Factor. This factor includes Comp Inc - Beginning Net Income, Pretax Income, Comprehensive Income-Parent, Special Items, Income Before Extra Items - Adj for Common Stock Equivalents - 12MM. Income is the company’s remaining revenues after paying all operating expenses and taxes. The higher the income, the probably better operating state.
5.2.3 Balance Sheet: Factor 3

The third factor is Implied Option Factor. This factor includes Implied Option EPS Diluted, Implied Option EPS Diluted Preliminary, Implied Option EPS Diluted 12MM, Implied Option Expense, Implied Option Expense Preliminary, Implied Option Expense - 12mm. The implied option mainly refers to the pre-redemption option, which is the risk of the interest rate that the borrower can repay the risk in advance. The existence of implied options is one of the important reasons for financial risk.

5.2.4 Balance Sheet: Factor 4

The third factor is Post Retirement Adjustment Factor. This factor includes Core Post Retirement Adjustment, Core Post Retirement Adjustment 12MM, Core Post Retirement Adjustment Diluted EPS Effect 12MM, Core Post Retirement Adjustment Diluted EPS Effect, Core Post Retirement Adjustment 12MM Diluted EPS Effect Preliminary. Retirement Adjustment will have effect on employees, which will influence operating of the company.
5.2.5 Balance Sheet: Factor 5

This factor is Pension Adjustment Factor. This factor includes Core Pension Interest Adjustment After-tax Preliminary, Core Pension Adjustment Preliminary, Core Pension Adjustment, Core Pension Adjustment Diluted EPS Effect 12MM, Core Pension Adjustment Diluted EPS Effect. The Pension Adjustment is a combination of all employee and employer pension credits for the year. The higher pension, the company have more funds to operate.
5.2.6 Balance Sheet: Factor 6

This factor is Earnings Per Share Factor. This factor includes Earnings Per Share (Diluted)-Excluding Extraordinary Items- 12 Months Moving, Earnings Per Share (Diluted)-Including Extraordinary Items, Core Pension w/o Interest Adjustment Diluted EPS Effect Preliminary. Earnings Per Share is the share of the company’s profits divided by its common stock. It indicates the profitability of a company. The higher the Earnings Per Share, the more profit a company obtains.
5.2.7 Balance Sheet: Factor 7

This factor is Discontinued Operations Factor. This factor includes Discontinued Operations, Extraordinary Items and Discontinued Operations. Discontinued operations is shown separately in the income statement, and investors can clearly know the profits and cash flows from the discontinued activities, especially when the company is merged, and it is better to analyze which assets are stripped or folded to understand how a company will make money in the future.
5.2.8 Balance Sheet: Factor 8

This factor is Preferred Dividends Factor. This factor includes Nonred Pfd Shares Outs (000) - Quarterly, Preferred/Preference Stock (Capital)-Total, Dividends-Preferred/Preference. A preferred dividend refers to the dividends accrued and paid by the company’s preferred shares. The higher the company’s earnings per share, the higher its profitability.

Figure 5.15: Balance Sheet: Factor 8

5.2.9 Balance Sheet: Factor 9

This factor is Non-Controlling Interests Factor. This factor includes Non-Controlling Interest - Redeemable - Balance Sheet, Non-Controlling Interests - Total - Balance Sheet. Non-Controlling Interests is not owned by the parent company, which owns a controlling stake (more than 50 per cent but less than 100 per cent) and merges the financial performance of the subsidiary into its own company. For investors, it is important for companies to provide transparency of non-controlled benefits, because this will enable them to better understand the impact of non-controlled benefits on the financial position, financial performance and cash flow of the company, and thus also understand the risks the company faced.
5.2.10 Balance Sheet: Factor 10

This factor is In Process R&D Factor. This factor includes In Process R&D, In Process R&D Expense After-tax. A company typically generates research and development costs as it seeks and creates new products or services. It is important for the company to maintain the forefront of innovation and invest heavily in product research and development. With the development of technology, these efforts enable the company to diversify and find new growth opportunities.

6 Features Selected Comparison

Top 10 features are selected in Financial Ratio and Balance Sheet according to the PCA(Algorithm 1) and FA(Algorithm 2). Then these features are com-
pared with the features selected by exploited Chi-Square and Genetic algorithm (GA). The feature overlaps of PCA (Algorithm 1), Chi-square and GA (Random Forest) are marked by orange and the feature overlaps of FA (Algorithm 2), Chi-square and GA (Random Forest) are green. The overlaps are the red letters with green background if the features selected by all four algorithms. Because the KMO value of feature set of Energy Sector in Balance Sheet is less than 0.6 (0.5) that is not suitable for Factor Analysis, there are no selected features in Energy Sector in Balance Sheet.

6.1 Financial Sector

Figure 6.1: Financial Sector Detailed Mapping in Financial Ratio
### Figure 6.2: Financial Sector Detailed Mapping in Balance Sheet

| Methods | Principal Component Analysis (Algorithm 1) | Chi-square Test |
|---------|---------------------------------------------|-----------------|
| Features | Noncontrolling Interest-Redeemable-Balance Sheet | Accumulated Other Comprehensive Income (Loss) |
|         | Deferred Taxes and Investment, Tax Credit | Current Assets - Other - Total |
|         | Common Shares Used to Calculate Earnings Per Share-12 Months Moving | Other Long term Assets |
|         | Cost of Goods Sold | Accum Other Comp Inc - Derivatives Unrealized Gain/Loss |
|         | Accounting Changes-Cumulative Effect | Accum Other Comp Inc - Other Adjustments |
|         | Stock Compensation Expense | Accum Other Comp Inc - Unreal G/L Ret for in Sec Assets |
|         | Preferred, ESOP Obligation-Non-Redeemable | Account Payable/Creditors - Trade |
|         | Property, Plant and Equipment (Total) - Net | Comp Inc - Currency Trans Adj |
|         | Core Pension, Adjustment Diluted EPS Effect | Non-Operating Income (Expense) - Total |
|         | Extraordinary Items | Accum Other Comp Inc - Min Pension Liability Adj |

### Figure 6.3: Financial Sector Coarse Mapping in Financial Ratio

| Methods | Principal Component Analysis (Algorithm 1) | Chi-square Test |
|---------|---------------------------------------------|-----------------|
| Features | Receivables / Turnover | Book/Market |
|         | Total Debt Equity | Enterprise Value Multiple |
|         | Long-term Debt/Invested Capital | Price/Operating Earnings (Basic, Excel EI) |
|         | Research and Development Sales | P/E (Diluted, Excel EI) |
|         | Capitalization Ratio | Price/Cash flow |
|         | Payables, Turnover | Net Profit Margin |
|         | P/E (Diluted, Excel EI) | Averting Expenses/Sales |
|         | Short-Term Debt/Total Debt | Labor Expenses/Sales |
|         | Total Debt/Total Assets | P/E (Diluted, Incl. EI) |
|         | Total Debt/Capital | Effective Tax Rate |

### Methods

- **Factor Analysis (Algorithm 2)**
  - GA (Support Vector Machine)
  - GA (Random Forest)
6.2 Energy Sector

Figure 6.5: Energy Sector Detailed Mapping in Financial Ratios

| Methods | Principal Component Analysis (Algorithm1) | Chi-square Test |
|---------|------------------------------------------|-----------------|
| Features | Price/Cash Flow | Book/Market |
|          | Receivables/turnover | Enterprise Value Multiple |
|          | P/E (Diluted, Excl. E1) | P/E (Diluted, Excl. E1) |
|          | Research and Development Sales | P/E (Diluted, Excl. E1) |
|          | Advertising Expenses/Sales | Price/Sales |
|          | Cash Flow/Margin | Receivables Turnover |
|          | Total Liabilities/Total Tangible Assets | Price/Cash Flow |
|          | Short/Long Term Debt/Total Debt | Operating Profit Margin After Depreciation |
|          | Common Equity/Invested Capital | Pre-tax Return on Total Earning Assets |

Methods |
---------|
Factor Analysis (Algorithm2) | GA (Support Vector Machine) |

Features |
Pre-tax Return on Total Earning Assets | Inventory/Current Assets |
Return on Capital Employed | Current Liabilities/Total Liabilities |
After tax Return on Average Common Equity | Long-term Debit/Total Liabilities |
Return on Equity | Total Debit/Total Assets |
Return on Assets | Cash Flow/Total Debt |
After tax Return on Invested Capital | Common Equity/Invested Capital |
Operating Profit Margin After Depreciation | Current Ratio |
Net Profit Margin | Cash Ratio |
Gross Profit/Total Assets | Payables Turnover |
Profit Before Depreciation/Current Liabilities | Total Debit/Capital |
Figure 6.6: Energy Sector Detailed Mapping in Balance Sheet

| Balance Sheet+Detailed Mapping | Chi-square Test |
|-------------------------------|-----------------|
| **Methods**                   |                 |
| Non-Operating Income/(Expense)-Total | Assets Level2 (Observable) |
| Other Long-term Assets | Current Assets - Other - Total |
| Gain/Loss on Sale/Core Earnings Adjusted/Afair tax. 12MM | Cash and Short Term Investments |
| Non-Current Assets - Total | Comp Inc - Minimum Pension Adj |
| Comp Inc-Derivative Gains/Losses | Deferred Revenue - Current |
| Derivative Liabilities, Long Term | Gain/Loss on Sale (Core Earnings Adjusted) Diluted EPS Effect 12MM |
| Pension Core Adjustment-12MM | Other Intangibles |
| Goodwill net. | Short-Term Investments - Total |
| Current Assets - Other - Total | Current Deferred Tax Liability |
| Income Before Extra Items- | Comp Inc - Securities Gains/Losses |
| Adj for Common Stock-Equivalents-12MM | |

| Features | |
|-----------|-----------------|
| **Factor Analysis(Algorithm2)** | GA(Support Vector Machine) |
| Capital Surplus/Share Premium Reserve | |
| Long-Term Debt - Total | |
| Inventories - Total | |
| Intangible Assets - Total | |
| Depreciation, Depletion and Amortization(Accumulated) | |
| Common/Ordinary Stock (Capital) | |
| Deferred Taxes and Investment Tax Credit | |
| Current Liabilities - Other - Total | |
| Assets - Other - Total | |
| Preferred/Preference Stock - Nonredeemable | |

Figure 6.7: Energy Sector Coarse Mapping in Financial Ratios

| Financial Ratio +Course Mapping | Chi-square Test |
|--------------------------------|-----------------|
| **Methods**                   |                 |
| P/E (Diluted, Excl. ED) | Book/Market |
| Book/Market |                 |
| Operating CF Current Liabilities | Price/Operating Earnings (Basic, Excl. ED) |
| P/E (Diluted, Incl. ED) | Research and Development/ Sales |
| Short-Term Debt/Total Debt | P/E (Diluted, Incl. ED) |
| Price/Book | Inventory Turnover |
| Return on Assets | Sales/Working Capital |
| Return on Equity | Price/Sales |
| Net Profit Margin | Inventory/Current Assets |
| Inventory/Current Assets | Long-term Debt/Book Equity |

| Features | |
|-----------|-----------------|
| **Factor Analysis(Algorithm2)** | GA(RandomForest) |
| Pre-tax,Return on Total Earning Assets | Labor Expenses/Sales |
| Return on Capital Employed | Total Liabilities/Total Tangible Assets |
| After-tax,Return on Average Common Equity | Short-Term Debt/Total Debt |
| Return on Equity | Advertising Expenses/Sales |
| Return on Assets | Total Debt/Total Assets |
| After-tax Return on Invested Capital | Total Debt/Total Assets |
| Operating Profit Margin After Depreciation | Asset Turnover |
| Net Profit Margin | Receivables Turnover |
| Gross Profit/Total Assets | Price/Sales |
| Profit Before Depreciation,Current Liabilities | Long-term Debt/Total Liabilities |
Figure 6.8: Energy Sector Coarse Mapping in Balance Sheet

| Balance Sheet+Coarse Mapping | Chi-square Test |
|-----------------------------|----------------|
| **Methods**                 | **Features**   |
| Core Post-Retirement Adjustment 12M Diluted EPS Effect Preliminary | Current Assets - Other - Total |
| Interest and Related Expense - Total | Accrued Other Comp Inc - Derivatives Unrealized Gain/Loss |
| Implied Option EPS Diluted Preliminary | Non-Current Assets - Total |
| Preferred ESOP Obligation...Non Redeemable | Cash and Short-Term Investments |
| Implied Option Expense Preliminary | Capital Surplus/Share Premium Reserve |
| Profit Income | Cost of Goods Sold |
| Implied Option EPS Diluted | Deferred Compensation |
| Current Assets-Other-Total | Comprehensive Income, Noncontrolling Interest |
| Cash and Short-Term Investments | Gain/Loss on Sale (Core Earnings Adjusted) After-tax 12MM |
| Carrying Value | Current Liabilities - Other - Total |

Methods: Factor Analysis (Algorithm 2) | GA (Random Forest)

Features: NA

6.3 Health Sector

Figure 6.9: Health Sector Detailed Mapping in Financial Ratios

| Financial Ratio +Detailed Mapping | Chi-square Test |
|----------------------------------|----------------|
| **Methods**                      | **Features**   |
| Pre-tax Return on Total Earning Assets | Book/Market |
| Free Cash Flow, Operating Cash Flow | Enterprise Value Multiple |
| After-tax Return on Average Common Equity | Sales/Invested Capital |
| Inventory/Current Assets | P/E (Diluted, Incl. EI) |
| Common Equity/Invested Capital | Price/Sales |
| Operating CF/Current Liabilities | Price/Operating Earnings (Basic, Excl. EI) |
| Sales/Stockholders Equity | P/E (Diluted, Excl. EI) |
| Sales/Working Capital | Price/Operating Earnings (Diluted, Excl. EI) |
| Price/Sales | Total Debt/Total Assets |
| Return on Assets | Price/Cash Flow |

Methods: Factor Analysis (Algorithm 2) | GA (Support Vector Machine)

Features: NA

57
### Figure 6.10: Health Sector Detailed Mapping in Balance Sheet

| Features | Principal Component Analysis (Algorithm 1) | Chi-square Test |
|----------|--------------------------------------------|-----------------|
| Core Post Retirement Adjustment - Diluted: EPS Effect Preliminary | Accumulated Other Comprehensive Income (Loss) | |
| In Process R&D Expense, After tax | Current Liabilities - Other - Total | |
| Core Post Retirement Adjustment: Diluted EPS, Effect 12MM | Accrued Other Comp Inc - Min Pension Liab Adj | |
| Debt in Current Liabilities | Carrying Value | |
| Current Liabilities - Other - Total | Accrued Other Comp Inc - Derivatives Unrealized Gain/Loss | |
| Earnings Per Share (Diluted): Excluding Extraordinary Items | Non-Current Assets - Total | |
| S.P. Core Earnings: Diluted, Preliminary | Comp Inc - Currency Trans Adj | |
| Interest and Related Expense, Total | Pension Core Adjustment 12Mm | |
| Total Long Term Investments | Goodwill (net) | |
| Core Post Retirement Adjustment Preliminary | Earnings Per Share (Diluted) - Including Extraordinary Items | |

| Methods | Factor Analysis (Algorithm 2) | GA (Support Vector Machine) |
|---------|-----------------------------|----------------------------|
| Current Liabilities - Other - Total | Inventories - Total | |
| Non Current Assets - Total | Property Plant and Equipment - Total (Net) | |
| Current Assets - Total | Interest and Related Expense - Total | |
| Operating Income, After Depreciation - Quarterly | Goodwill (net) | |
| Depreciation and Amortization - Total | Selling, General and Administrative Expenses | |
| Long Term Liabilities - Total | Unadjusted Retained Earnings | |
| Current Assets - Other - Total | Noncontrolling Interests - Total - Balance Sheet | |
| Common Shares Used to Calculate Earnings Per Share - 12 Months Moving | Long-Term Debt - Total | |
| Property Plant and Equipment - Total (Net) | Retained Earnings | |
| Cash and Short-Term Investments | Assets - Other - Total | |

### Figure 6.11: Health Sector Coarse Mapping in Financial Ratios

| Features | Principal Component Analysis (Algorithm 1) | Chi-square Test |
|----------|--------------------------------------------|-----------------|
| P/E (Diluted, Excl. E) | Book/Market | |
| Cash Flow/Total Debt | Enterprise Value Multiple | |
| Pre-tax Return on Net/Operating Assets | Price/Operating Earnings (Diluted, Excl. E) | |
| Gross Profit/Total Assets | Price/Operating Earnings (Basic, Excl. E) | |
| Cash Balance/Total Liabilities | Price/Sales | |
| Total Debt/Total Assets | P/E (Diluted, Excl. E) | |
| Effective Tax Rate | After-tax Return on Average Common Equity | |
| Profit Before Depreciation/Current Liabilities | Price/Cash flow | |
| Price/Book | Net Profit Margin | |
| Total Liabilities/Total Tangible Assets | Effective Tax Rate | |

| Methods | Factor Analysis (Algorithm 2) | GA (Random Forest) |
|---------|-----------------------------|-------------------|
| Long-term Debt/Invested Capital | Gross Profit Margin | |
| Total Debt/Invested Capital | Effective Tax Rate | |
| Common Equity/Invested Capital | Cash Flow/Total Debt | |
| Total Debt/Total Assets | Research and Development/Sales | |
| Total Debt/Capital | Total Liabilities/Total Tangible Assets | |
| Long-term Debt Book Equity | Advertising Expenses/Sales | |
| Total Debt/Total Assets | Dividend Payout Ratio | |
| Long-term Debt/Total Liabilities | Shillers Cyclically Adjusted P/E Ratio | |
| Return on Equity | Inventory Turnover | |
| After-tax Return on Invested Capital | Receivables Turnover |
### 6.4 Consumer Sector

Figure 6.13: Consumer Sector Detailed Mapping in Financial Ratios
Figure 6.14: Consumer Sector Detailed Mapping in Balance Sheet

### Balance Sheet Detailed Mapping

| Methods                  | Principal Component Analysis (Algorithm1) | Chi-square Test       |
|--------------------------|------------------------------------------|-----------------------|
| Operating Income Before Depreciation - Quarterly | Accumulated Other Comprehensive Income (Loss) |
| Non-Operating Income/Expense - Total | Current Assets - Total                   |
| Core Pension Adjustment | Other Long-Term Assets                   |
| Core Post Retirement Adjustment 12MM Diluted EPS Effect Preliminary | Assets - Other - Total |
| Features                 |                                           |                       |
| Common Shares Used to Calculate Earnings Per Share - 12 Months Moving | Capital Surplus/Share Premium Reserve |
| Dilution Adjustment | Total Shares Repurchased - Quarter        |
| Accum Other Comp Inc...Derivatives, Unrealized Gain/Loss | Discontinued Operations |
| Core Post Retirement Adjustment Preliminary | Common Stock Equivalents - Dollar Savings |
| Accounting Changes...Cumulative Effect | Current Assets - Other - Total |

| Methods                  | Factor Analysis (Algorithm2) | GA (Support Vector Machine) |
|--------------------------|----------------------------|-----------------------------|
| Interest and Related Expense - Total | Carrying Value              |
| Cost of Goods Sold | Property Plant and Equipment - Total (Net) |
| Depreciation and Amortization - Total | Unadjusted Retained Earnings |
| Long Term Debt - Total | Inventories - Total         |
| Features                 | Property Plant and Equipment - Total (Net) | Treasury Stock - Total (All Capital) |
| Operating Income Before Depreciation - Quarterly | Account Payable/Creditors - Trade |
| Liabilities - Other | Common Shares Issued        |
| Current Liabilities - Other - Total | Long-Term Liabilities (Total) |
| Operating Income After Depreciation - Quarterly | Income Taxes Payable |
| Assets - Other - Total | Capital Surplus/Share Premium Reserve |

Figure 6.15: Consumer Sector Coarse Mapping in Financial Ratios

### Financial Ratio Coarse Mapping

| Methods                  | Principal Component Analysis (Algorithm1) | Chi-square Test       |
|--------------------------|------------------------------------------|-----------------------|
| Trailing P.E to Growth (PEG) ratio | Book/Market |
| Common Equity Invested Capital | Enterprise Value Multiple |
| Pre-tax Return on Total Earning Assets | Price Operating Earnings (Basic, Excl. EI) |
| Return on Assets | P/E (Diluted, Excl. EI) |
| Gross Profit Margin | P/E (Diluted, Incl. EI) |
| Long-term Debt/Invested Capital | Long-term Debt/Total Liabilities |
| Net Profit Margin | Effective Tax Rate |
| Effective Tax Rate | Price/Cash Flow |
| Long-term Debt/Total Liabilities | After-tax Return on Invested Capital |
| Short Term Debt/Total Debt | Inventory/Current Assets |
| Features                 |                                           |                       |
| Pre-tax Return on Total Earning Assets | Advertising Expenses/Sales |
| Return on Assets | Gross Profit Margin |
| Profit Before Depreciation, Current Liabilities | Effective Tax Rate |
| Accruals, Average Assets | Cash Flow/Total Debt |
| Return on Capital/Employee | Total Liabilities/Total Tangible Assets |
| Total Debt/Total Assets | Research and Development/Sales |
| Total Debt/Total Assets, 1 | Dividend Payout Ratio |
| Total Debt/Total Capital | Receivables Turnover |
| Cash Ratio | Asset Turnover |
| Quick Ratio (Acid Test) | Operating Profit Margin Before |
From above figures, we selected the features by Algorithm 1 or Algorithm 2, Chi-Square and Genetic Algorithm at the same time. We summarised features according to data sets, Financial Ratio Detailed Mapping, Financial Ratio Coarse Mapping, Balance Sheet Detailed Mapping, Balance Sheet Coarse Mapping. We displayed the feature selected at the same time in two or more sectors in every data set. In Financial Ratio Detailed Mapping in four sectors, the features appeared in two or more sectors were Receivables / Turnover.

Algorithm 1 in Financial Ratio Coarse Mapping in four sectors, the features appeared in two or more sectors were P/E (Diluted, Excl, EI).

Algorithm 2 in Financial Ratio Detailed Mapping in four sectors, the features appeared in two or more sectors were Total Debt / Total Assets, Total Debt / Total Liabilities.

Algorithm 2 in Financial Ratio Coarse Mapping in four sectors, there was no features that appeared in two or more sectors.

Algorithm 1 in Balance Sheet Detailed Mapping in four sectors, the features appeared in two or more sectors were Assets - Other - Total.

Algorithm 1 in Balance Sheet Coarse Mapping in four sectors, the features appeared in two or more sectors were Carrying Value.

Algorithm 2 in Balance Sheet Detailed Mapping in four sectors, the features appeared in two or more sectors were Property Plant and Equipment - Total (Net).

Algorithm 2 in Balance Sheet Coarse Mapping in four sectors, the features appeared in two or more sectors were Common Shares Used to Calculate...
From above results, we know that Algorithm 2 selects more overlapping features compared with Algorithm 1. This confirms Algorithm 2 has more commonality with other feature selection algorithms. And also in previous section, the Algorithm 2 needed lesser number of features when arrived the same accuracy in comparison to Algorithm 1. Algorithm 2 not only can improve the computational efficiency, but also can summarize the similar features together to explain one common factors. Therefore, Algorithm 2 are better to select features compared with Algorithm 1.
References

[1] Balance Sheet
https://www.investopedia.com/terms/b/balancesheet.asp

[2] Bandalos D L, Finney S J. Factor analysis: Exploratory and confirmatory[M]//The reviewer’s guide to quantitative methods in the social sciences. Routledge, 2018: 98-122.

[3] Credit Rating
https://www.investopedia.com/terms/c/creditrating.asp

[4] Exploratory Factor Analysis
https://www.statisticssolutions.com/factor-analysis-sem-exploratory-factor-analysis/

[5] Factor Analysis
https://en.wikipedia.org/wiki/Factor_analysis

[6] Financial Ratios Based on the Balance Sheet
https://www.accountingcoach.com/financial-ratios/explanation/2

[7] Florescu I, Alves T, Bozdog D, et al. Using SHIFT a laboratory environment to test changes to financial markets[C]//2020 Fall Central Sectional Meeting. AMS, 2020.

[8] Golbayani P, Florescu I, Chatterjee R. A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees[J]. The North American Journal of Economics and Finance, 2020, 54: 101251.

[9] Golbayani P, Wang D, Florescu I. Application of Deep Neural Networks to assess corporate Credit Rating[J]. arXiv preprint arXiv:2003.02334 2020.

[10] Hira Z M, Gillies D F. A review of feature selection and feature extraction methods applied on microarray data[J].

[11] Principal component analysis
https://en.wikipedia.org/wiki/Principal_component_analysis

[12] Wang D, Wang T, Florescu I. Is Image Encoding Beneficial for Deep Learning in Finance? An Analysis of Image Encoding Methods for the Application of Convolutional Neural Networks in Finance[J]. arXiv preprint arXiv:2010.08698 2020.

[13] Yu C L, Sheu C F. EFAshiny: An User-Friendly Shiny Application for Exploratory Factor Analysis[J]. J. Open Source Software, 2018, 3(22): 567. Advances in bioinformatics, 2015, 2015.

[14] Zhou Y, Han L, Wang D, et al. A moment-based criterion for determining the number of components in a normal mixture model[J]. Journal of Systems Engineering and Electronics, 2017, 28(4): 801-809.
Appendices

Top 10 features selected by Algorithm 1 and Algorithm 2 and Feature Classification

A  Top 10 features selected by Algorithm 1

A.1  Financial Sector

A.1.1  Financial Sector:Data1: Financial Ratio + Detailed Mapping

Receivables Turnover
Total Debt/Equity
Payables Turnover
Capitalization Ratio
Operating Profit Margin After Depreciation
Total Debt/Total Assets 1
Long-term Debt/Invested Capital
Research and Development/Sales
Total Liabilities/Total Tangible Assets
Book/Market

A.1.2  Financial Sector:Data2:Balance Sheet + Detailed Mapping

Noncontrolling Interest - Redeemable - Balance Sheet
Deferred Taxes and Investment Tax Credit
Common Shares Used to Calculate Earnings Per Share - 12 Months Moving
Cost of Goods Sold
Accounting Changes - Cumulative Effect
Stock Compensation Expense
Preferred ESOP Obligation - Non-Redeemable
Property Plant and Equipment - Total (Net)
Core Pension Adjustment Diluted EPS Effect
Extraordinary Items

A.1.3  Financial Sector:Data3:Financial Ratio + Coarse Mapping

Receivables Turnover
Total Debt/Equity
Long-term Debt/Invested Capital
Research and Development/Sales
Capitalization Ratio
Payables Turnover
P/E (Diluted, Excl. EI)
Short-Term Debt/Total Debt
Total Debt/Total Assets
Total Debt/Capital

A.1.4 Financial Sector: Data 4: Balance Sheet + Coarse Mapping
Core Post Retirement Adjustment Diluted EPS Effect 12MM
Interest and Related Expense-Total
Implied Option EPS Diluted Preliminary
Preferred ESOP Obligation-Non.Redeemable
Implied Option Expense Preliminary
Pretax Income
Implied Option EPS Diluted
Current Assets-Other-Total
Cash and Short Term Investments
Carrying Value

A.2 Energy Sector
A.2.1 Energy Sector: Data 1: Financial Ratio + Detailed Mapping
Price/Cash flow
Receivables Turnover
P.E(Diluted, Excl, EI)
Research and Development/Sales
Advertising Expenses/Sales
Cash Flow Margin
Total Liabilities/Total Tangible Assets
Dividend Payout Ratio
Short-Term Debt/Total Debt
Common Equity/Invested Capital

A.2.2 Energy Sector: Data 2: Balance Sheet + Detailed Mapping
Non-Operating Income (Expense) - Total
Other Long-term Assets
Gain Loss on Sale-Core Earnings Adjusted After-tax 12MM
Non-Current Assets-Total
Comp Inc-Derivative Gains Losses
Derivative Liabilities Long-Term
Pension Core Adjustment-12mm
Goodwill(net)
Current Assets-Other-Total

65
A.2.3 Energy Sector : Data 3: Financial Ratio + Coarse Mapping

- P/E (Diluted, Excl. EI)
- Book/Market
- Operating CF/Current Liabilities
- P/E (Diluted, Incl. EI)
- Short-Term Debt/Total Debt
- Price/Book
- Return on Assets
- Return on Equity
- Net Profit Margin
- Inventory/Current Assets

A.2.4 Energy Sector : Data 4: Balance Sheet + Coarse Mapping

- Core Post Retirement Adjustment 12MM Diluted EPS Effect Preliminary
- Interest and Related Expense-Total
- Implied Option EPS Diluted Preliminary
- Preferred ESOP Obligation-Non Redeemable
- Implied Option Expense Preliminary
- Pretax Income
- Implied Option EPS Diluted
- Current Assets-Other-Total
- Cash and Short-Term Investments
- Carrying Value

A.3 Health Sector

A.3.1 Health Sector : Data 1: Financial Ratio + Detailed Mapping

- Pre-tax Return on Total Earning Assets
- Free Cash Flow Operating Cash Flow
- After-tax Return on Average Common Equity
- Inventory/Current Assets
- Common Equity /Invested Capital
- Operating CF/Current Liabilities
- Sales/Stockholders Equity Sales/Working Capital
- Price/Sales
- Return on Assets
A.3.2 Health Sector : Data2: Balance Sheet + Detailed Mapping

Core Post Retirement Adjustment 12MM Diluted EPS Effect Preliminary
In Process RD Expense Aftertax
Core Post Retirement Adjustment Diluted EPS Effect 12MM
Debt in Current Liabilities
Current Liabilities-Other-Total
Earnings Per Share (Diluted)-Excluding Extraordinary items
SPCore Earnings EPS Diluted Preliminary
Interest and Related Expense Total
Total Long-term Investments
Core Post Retirement Adjustment Preliminary

A.3.3 Health Sector : Data3: Financial Ratio + Coarse Mapping

P/E (Diluted, Excl, EI) Cash Flow/Total Debt
Pretax return on Net/Operating Assets
Gross Profit/Total Assets
Cash Balance/Total Liabilities
Total Debt/Total Assets
Effective Tax Rate
Profit Before Depreciation/Current Liabilities
Price/Book
Total Liabilities/Total Tangible Assets

A.3.4 Health Sector : Data 4: Balance Sheet + Coarse Mapping

Preferred ESOP Obligation - Non-Redeemable
In Process R&D Expense After-tax
Depreciation, Depletion and Amortization (Accumulated)
Treasury Stock - Total (All Capital)
Non-controlling Interests - Total Balance Sheet
Non-Current Assets - Total
Core Post Retirement Adjustment Diluted EPS Effect 12MM
Core Post Retirement Adjustment 12MM Diluted EPS Effect Preliminary
Implied Option Expense
Core Post Retirement Adjustment Preliminary
A.4 Consumer Sector

A.4.1 Consumer Sector : Data1: Financial Ratio + Detailed Mapping

P/E (Diluted, Incl. EI) Common Equity/Invested Capital Long-term Debt/Invested Capital Pre-tax Return on Total Earning Assets Net Profit Margin Gross Profit Margin Effective Tax Rate Current Liabilities/Total Liabilities Total Debt/Total Assets Profit Before Depreciation/Current Liabilities

A.4.2 Consumer Sector : Data2: Balance Sheet + Detailed Mapping

Operating Income Before Depreciation-Quarterly Non Operating Income(Expense)-Total Core Pension Adjustment Core Post Retirement Adjustment 12MM Diluted EPS Effect Preliminary Common Shares Used to Calculate Earnings Per Share - 12 Months Moving Dilution Adjustment Accum Other Comp Inc - Derivatives Unrealized Gain/Loss Core Post Retirement Adjustment Preliminary Accounting Changes-Cumulative Effect

A.4.3 Consumer Sector : Data3 : Financial Ratio + Coarse Mapping

Trailing P/E to Growth (PEG) ratio Common Equity/Invested Capital Pre-tax Return on Total Earning Assets Return on Assets Gross Profit Margin Long-term Debt/Invested Capital Net Profit Margin Effective Tax Rate Long-term Debt/Total Liabilities Short Term Debt/Total Debt

A.4.4 Consumer Sector : Data 4 : Balance Sheet + Coarse Mapping

Operating Income Before Depreciation-Quarterly Non-Operating Income(Expense)-Total Core Pension Adjustment Core Post Retirement Adjustment 12MM Diluted EPS Effect Preliminary Common Shares Used to Calculate Earnings Per Share-12.Months.Moving Dilution Adjustment Accum Other Comp Inc - Derivatives Unrealized Gain/Loss Core Post Retirement Adjustment Preliminary
B  Top 10 features selected by Algorithm 2

B.1  Financial Sector

B.1.1  Financial Sector: Data1: Financial Ratio + Detailed Mapping

- Total Debt/Total Assets
- Total Debt/EBITDA
- Long-term Debt/Total Liabilities
- Cash Flow/Total Debt
- Long-term Debt/Book Equity
- Total Debt/Capital
- Total Debt/Equity
- Asset Turnover
- Receivables Turnover
- Payables Turnover

B.1.2  Financial Sector: Data2: Balance Sheet + Detailed Mapping

- Account Payable Creditors-Trade
- Long-Term Debt-Total
- Capital Surplus Share Premium Reserve
- Cash and Short-Term Investments
- Goodwill(net)
- Non-Operating Income (Expense) - Total
- Common Shares Used to Calculate Earnings Per Share - 12 Months Moving
- Preferred/Preference Stock - Nonredeemable
- Sales Turnover (Net)
- Depreciation and Amortization-Total

B.1.3  Financial Sector: Data3 : Financial Ratio + Coarse Mapping

- Total Debt /Total Assets 1
- Asset Turnover
- Return on Assets
- Total Debt/Equity
- Cash Flow/Total Debt
- Long-term Debt/Total Liabilities
- Total Debt/Capital
- Receivables Turnover
- Gross.Profit/Total Assets
Capitalization Ratio

B.1.4 **Financial Sector: Data 4 : Balance Sheet + Coarse Mapping**
- Account Payable/Creditors-Trade
- Long-Term Debt-Total
- Capital Surplus/Share Premium Reserve
- Cash and Short-Term Investments
- Goodwill(net)
- Non-Operating Income (Expense)-Total
- Common Shares Used to Calculate Earnings Per Share-12 Months Moving
- Sales Turnover (Net)
- Depreciation and Amortization -Total
- Retained Earnings

B.2 **Energy Sector**

B.2.1 **Energy Sector : Data1: Financial Ratio + Detailed Mapping**
- Pre-tax Return on Total Earning Assets
- Return on Capital Employed
- After-tax Return on Average Common Equity
- Return on Equity
- Return on Assets
- After-tax Return on Invested Capital
- Operating Profit Margin After Depreciation
- Net Profit Margin
- Gross Profit/Total Assets
- Profit Before Depreciation/Current Liabilities

B.2.2 **Energy Sector : Data3 : Financial Ratio + Coarse Mapping**
- Pre-tax Return on Total Earning Assets
- Return on Capital Employed
- After-tax Return on Average Common Equity
- Return on Equity
- Return on Assets
- After-tax Return on Invested Capital
- Operating Profit Margin After Depreciation
- Net Profit Margin
- Gross Profit/Total Assets
- Profit Before Depreciation /Current Liabilities
B.3 Health Sector

B.3.1 Health Sector : Data1: Financial Ratio + Detailed Mapping

- Long-term Debt/Invested Capital
- Common Equity/Invested Capital
- Total Debt/Invested Capital
- Total Debt/Total Assets
- Total Debt/Capital
- Long-term Debt/Book Equity
- Total Debt/Total Assets1
- Long-term Debt/Total Liabilities
- Return on Equity
- After-tax Return on Invested Capital

B.3.2 Health Sector : Data2: Balance Sheet + Detailed Mapping

- Core Post Retirement Adjustment 12MM Diluted EPS Effect Preliminary
- In Process R&D Expense After-tax
- Core Post Retirement Adjustment Diluted EPS Effect 12MM
- Debtin Current Liabilities
- Current Liabilities-Other-Total
- Earnings Per Share(Diluted)-Excluding Extraordinaryitems
- SP Core Earnings EPS Diluted Preliminary
- Interest and Related Expense Total
- Total Long-term Investments
- Core Post Retirement Adjustment Preliminary

B.3.3 Health Sector : Data3: Financial Ratio + Coarse Mapping

- Long-term Debt/Invested Capital
- Total Debt/Invested Capital
- Common Equity/Invested Capital
- Total Debt/Total Assets
- Total Debt/Capital
- Long-term Debt/Book Equity
- Total Debt/Total Assets1
- Long-term Debt/Total Liabilities
- Return on Equity
- After-tax Return on Invested Capital

B.3.4 Health Sector : Data 4 : Balance Sheet + Coarse Mapping

- Non-Current Assets-Total
- Current Liabilities-Other-Total
Long-Term Liabilities-Total
Current Assets-Total
Depreciation and Amortization-Total
Operating Income After Depreciation Quarterly
Common Shares Used to Calculate Earnings Per Share-12 Months Moving
Current Assets-Other-Total
Property Plant and Equipment-Total(Net)
Research and Development Expense

B.4 Consumer Sector

B.4.1 Consumer Sector : Data1: Financial Ratio + Detailed Mapping

Pre-tax Return on Total Earning Assets
Return on Assets
Profit Before Depreciation Current Liabilities
Accruals Average Assets
Return on Capital Employed
Total Debt/Total Assets
Total Debt/Total Assets 1
Total Debt/Capital
Cash Ratio
Quick Ratio (Acid) Test

B.4.2 Consumer Sector : Data2: Balance Sheet + Detailed Mapping

Interest and Related Expense-Total
Cost of Goods Sold
Depreciation and Amortization-Total
Long-Term Debt-Total
Property Plant and Equipment - Total (Net)
Operating Income Before Depreciation-Quarterly
Liabilities-Other
Current Liabilities-Other-Total
Operating Income After Depreciation-Quarterly
Assets - Other - Total

B.4.3 Consumer Sector : Data3: Financial Ratio + Coarse Mapping

Pre-tax Return on Total Earning Assets
Return on Assets
Profit Before Depreciation Current Liabilities
Accruals Average Assets
Return on Capital/Employed
Total Debt/Total Assets
Total Debt/Total Assets 1
Total Debt/Capital
Cash Ratio
Quick Ratio(Quid)Test

B.4.4 Consumer Sector :Data 4 :Balance Sheet + Coarse Mapping

Special Items
Pretax Income
Comp Inc(Beginning)Net Income
Redeem Pfd Shares Outs (000)
Earnings Per Share(Diluted)-Including Extraordinary Items 1
Earnings Per Share(Diluted)from Operations
Income Before Extra Items-Adj for Common Stock Equivalents-12MM
Pension Core Adjustment-12mm
Core Pension Adjustment Diluted EPS Effect 12MM
Core Pension Adjustment Preliminary

C Feature Classification

C.1 Financial Sector

C.1.1 Financial Sector:Data 1:Financial Ratio Detailed Mapping

Figure C.1: Financial Sector:Data 1:Financial Ratio Detailed Mapping

(a) Factor 1  (b) Factor 2
C.1.2 Financial Sector: Data 2: Financial Ratio coarse Mapping

Figure C.2: Financial Sector: Data 2: Financial Ratio coarse Mapping

(a) Factor 1
(b) Factor 2
(c) Factor 3
(d) Factor 4
(e) Factor 5
(f) Factor 6
C.1.3 Financial Sector: Data 3: Balance Sheet Detailed Mapping

Figure C.3: Financial Sector: Data 3: Balance Sheet Detailed Mapping
C.1.4 Financial Sector: Data 4: Balance Sheet Coarse Mapping

Figure C.4: Financial Sector: Data 4: Balance Sheet Coarse Mapping

(a) Factor 1  
(b) Factor 2
C.2 Energy Sector

C.2.1 Energy Sector: Data 1: Financial Ratio Detailed Mapping

Figure C.5: Energy Sector: Data 1: Financial Ratio Detailed Mapping
C.2.2 Energy Sector: Data 2: Financial Ratio Coarse Mapping

Figure C.6: Energy Sector: Data 1: Financial Ratio Coarse Mapping

(a) Factor 1
(b) Factor 2
(c) Factor 3
(d) Factor 4
(e) Factor 5
(f) Factor 6
(g) Factor 7
(h) Factor 8
C.3 Health Sector

C.3.1 Health Sector: Data 1: Financial Ratio Detailed Mapping

Figure C.7: Health Sector: Data 1: Financial Ratio Detailed Mapping

(a) Factor 1

(b) Factor 2

(c) Factor 3

(d) Factor 4

(e) Factor 5

(f) Factor 6

(g) Factor 7

(h) Factor 8
C.3.2 Health Sector: Data 2: Financial Ratio Coarse Mapping

Figure C.8: Health Sector: Data 2: Financial Ratio Coarse Mapping
C.3.3 Health Sector: Data 3: Balance Sheet Detailed Mapping

Figure C.9: Health Sector: Data 3: Balance Sheet Detailed Mapping

(a) Factor 1  (b) Factor 2

(c) Factor 3  (d) Factor 4
C.3.4 Health Sector: Data 4: Balance Sheet Coarse Mapping

Figure C.10: Health Sector: Data 4: Balance Sheet Coarse Mapping

(a) Factor 1

(b) Factor 2
C.4 Consumer Sector

C.4.1 Consumer Sector: Data 1: Financial Ratio Detailed Mapping

Figure C.11: Consumer Sector: Data 1: Financial Ratio Detailed Mapping

(a) Factor 1

(b) Factor 2

(c) Factor 3

(d) Factor 4

(e) Factor 5

(f) Factor 6

(g) Factor 7

(h) Factor 8
C.4.2 Consumer Sector: Data 2: Financial Ratio Coarse Mapping

Figure C.12: Consumer Sector: Data 2: Financial Ratio Coarse Mapping
C.4.3 Consumer Sector: Data 3: Balance Sheet Detailed Mapping

Figure C.13: Consumer Sector: Data 3: Balance Sheet Detailed Mapping
C.4.4 Consumer Sector: Data 4: Balance Sheet Coarse Mapping

Figure C.14: Consumer Sector: Data 4: Balance Sheet Coarse Mapping

(a) Factor 1
(b) Factor 2
(c) Factor 3
(d) Factor 4
(e) Factor 5
(f) Factor 6
(g) Factor 7
(h) Factor 8
(i) Factor 9

(j) Factor 10

(k) Factor 11

(l) Factor 12