2018AIML600

# Chi Square Analysis

## Table 1: CHI Square Table

|  |  |  |  |
| --- | --- | --- | --- |
| Sr No | Features | CHI SCORE\* | DF |
| 1 | RI | 13.0 | 10 |
| 2 | NA | 780.58 | 705 |
| 3 | MG | 575.1 | 465 |
| 4 | AL | 670.97 | 585 |
| 5 | SI | 672.11 | 610 |
| 6 | K | 458.26 | 315 |
| 7 | CA | 816.7 | 670 |
| 8 | BA | 240.88 | 165 |
| 9 | FE | 118.9 | 150 |

\* Values are rounded off up to 2 decimal places

## Table 2: P-Value Table (From Most Significant Feature to Least Significant Feature)

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No | Features | P-Value | Significance |
| 1 | K | 2.2863e-07 | << 0.05 |
| 2 | CA | 8.31326e-05 | << 0.05 |
| 3 | BA | 0.0001065 | << 0.05 |
| 4 | MG | 0003615 | < 0.05 |
| 5 | AL | 0.00778 | < 0.05 |
| 6 | NA | 0.0248 | < 0.05 |
| 7 | SI | 0.04092 | < 0.05 |
| 8 | RI | 0.22367 | > 0.05 |
| 9 | FE | 0.9712 | > 0.05 |

# Information Gain Analysis

Features are ordered in table with Most Significant Feature to Least Significant Feature.

|  |  |  |
| --- | --- | --- |
| Sr. No | Features | Information Gain |
| 1 | AL | 0.601 |
| 2 | MG | 0.586 |
| 3 | K | 0.468 |
| 4 | CA | 0.436 |
| 5 | BA | 0.42 |
| 6 | NA | 0.354 |
| 7 | RI | 0.256 |
| 8 | SE | 0.129 |
| 9 | FE | 0.08 |

# Wrapper Method Analysis

There are 6 features selected as below (Algorithm used LinearSVC as classifier)

Selected 6 features Index [1, 2, 3, 5, 7, 8]

Selected features ['NA', 'MG', 'AL', 'K', 'BA', 'FE']

RMSE Score with selected features: 1.219