The data read in, clean (Holland and unemployed). Then partitioned and modeled. GLM, lasso (variations), RF compared by ROC. Cutoff found. Misclassification on testing. Final with all data.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters and Estimates | | | | | Intercept | | | | | |
| -3.19 | | | | | |
| Marital Status | | Occupation | | | | | | | | Sex |
| Married-Civ-Spouse | Never Married | Exec-managerial | Farming-fishing | Handlers-cleaners | | Other-service | Prof-specialty | Sales | Tech-support | Male |
| 1.79 | -0.44 | 0.84 | -0.94 | -0.63 | | -0.86 | 0.61 | 0.35 | 0.62 | 0.80 |
| Relationship | | Class | | | | Age | Education | Capital | | Hours |
| Own-child | Wife | Federal-gov | Self-emp-inc | Self-emp-  not-inc | | Years | Number | Loss | Gain | Per week |
| -0.99 | 1.17 | 0.57 | 0.20 | -0.46 | | 0.32 | 0.72 | 0.26 | 2.35 | 0.38 |

Chart showing an important relationship

