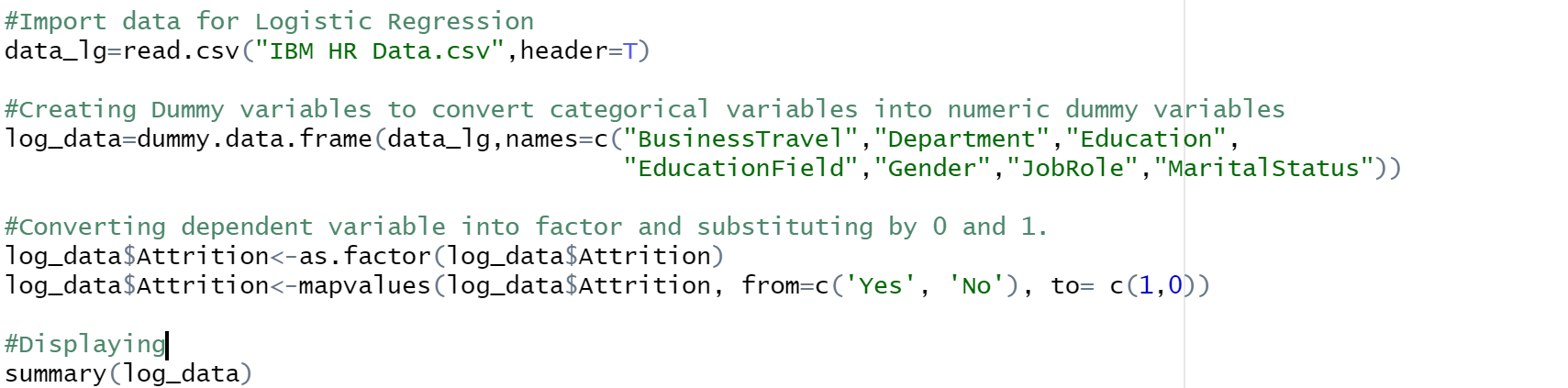
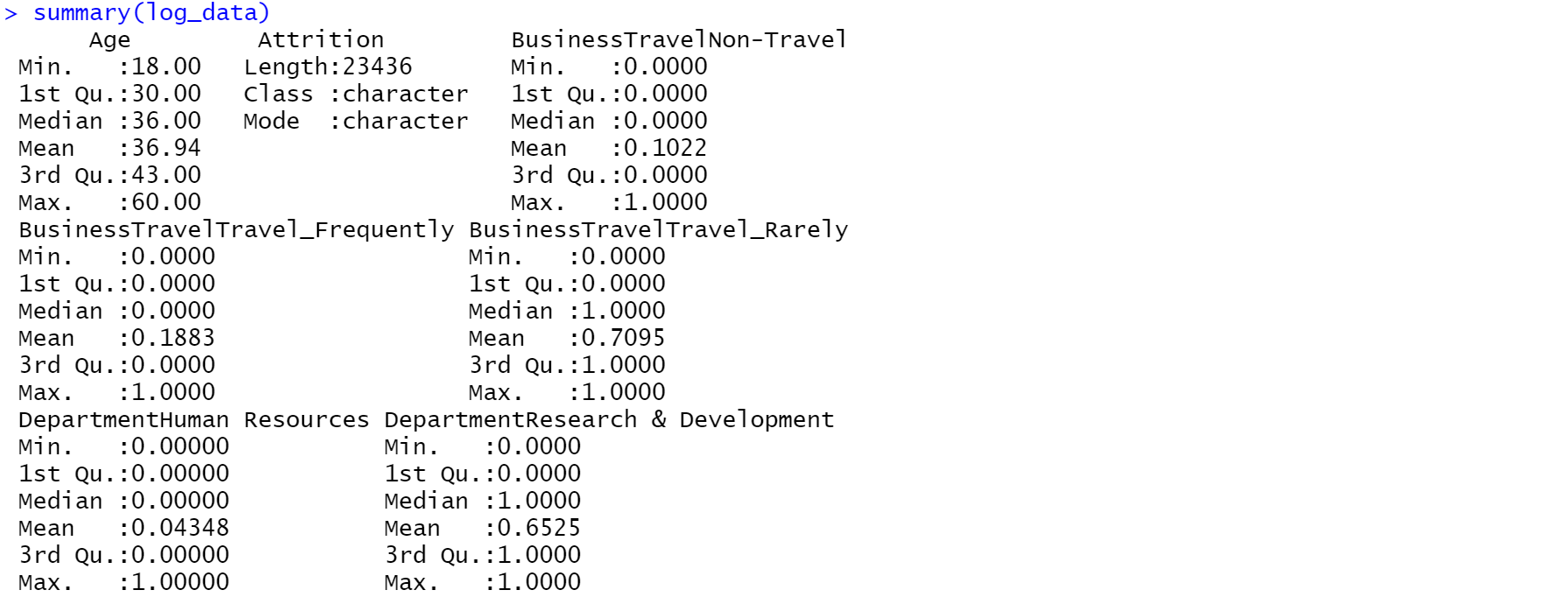
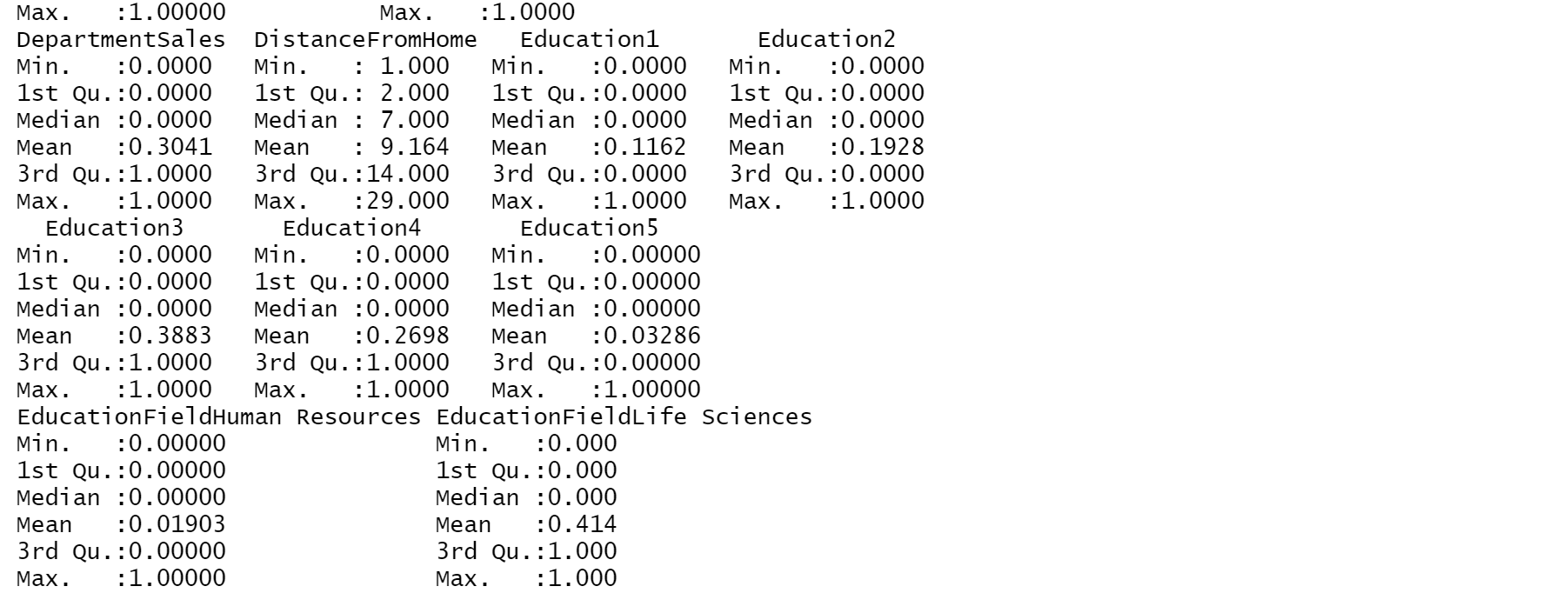
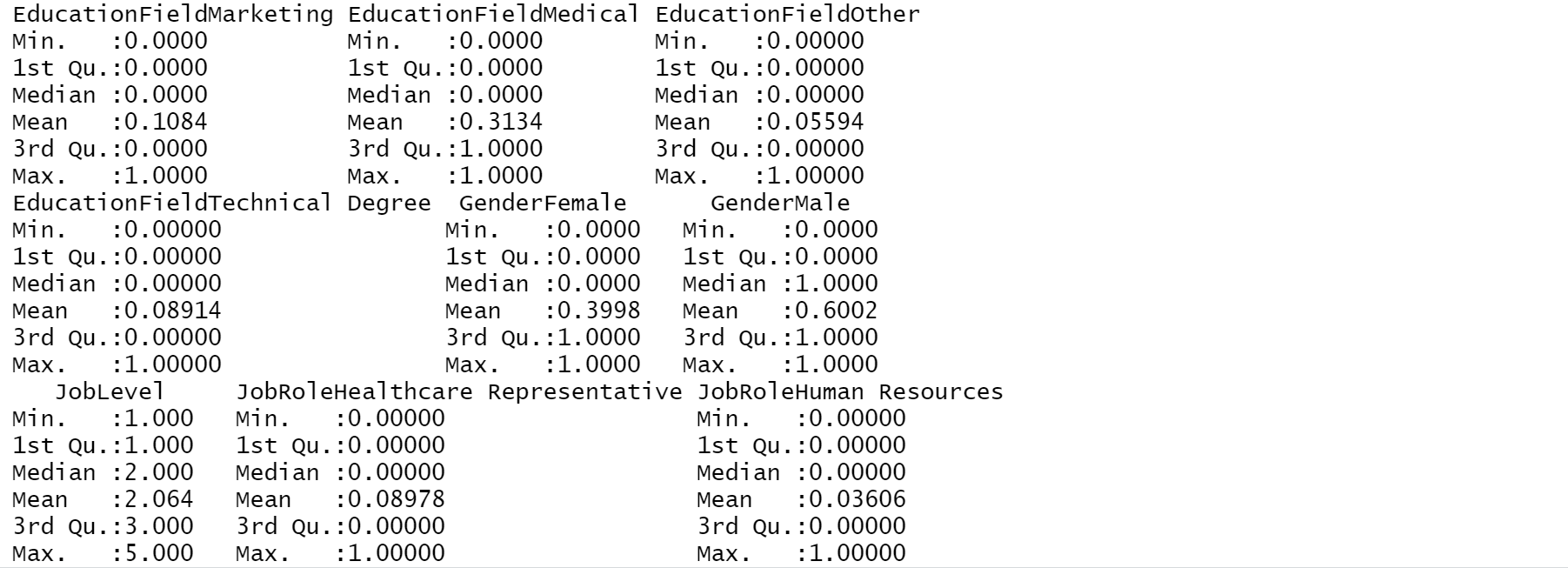
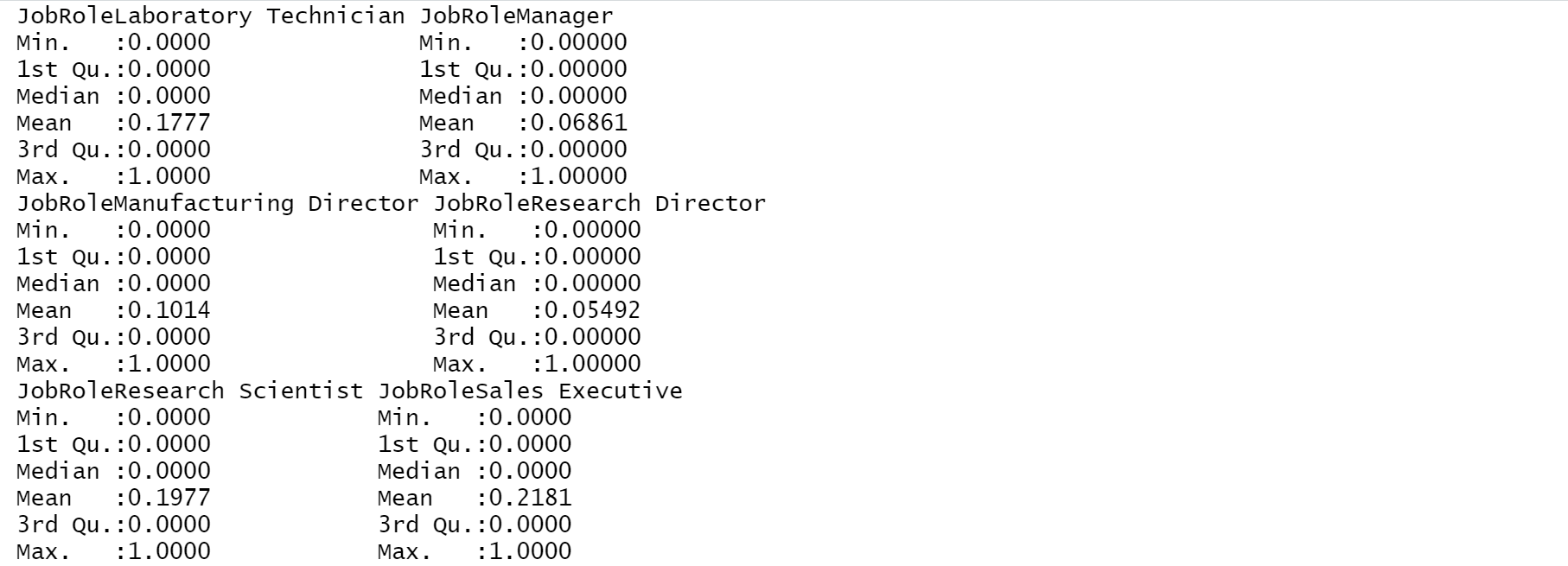
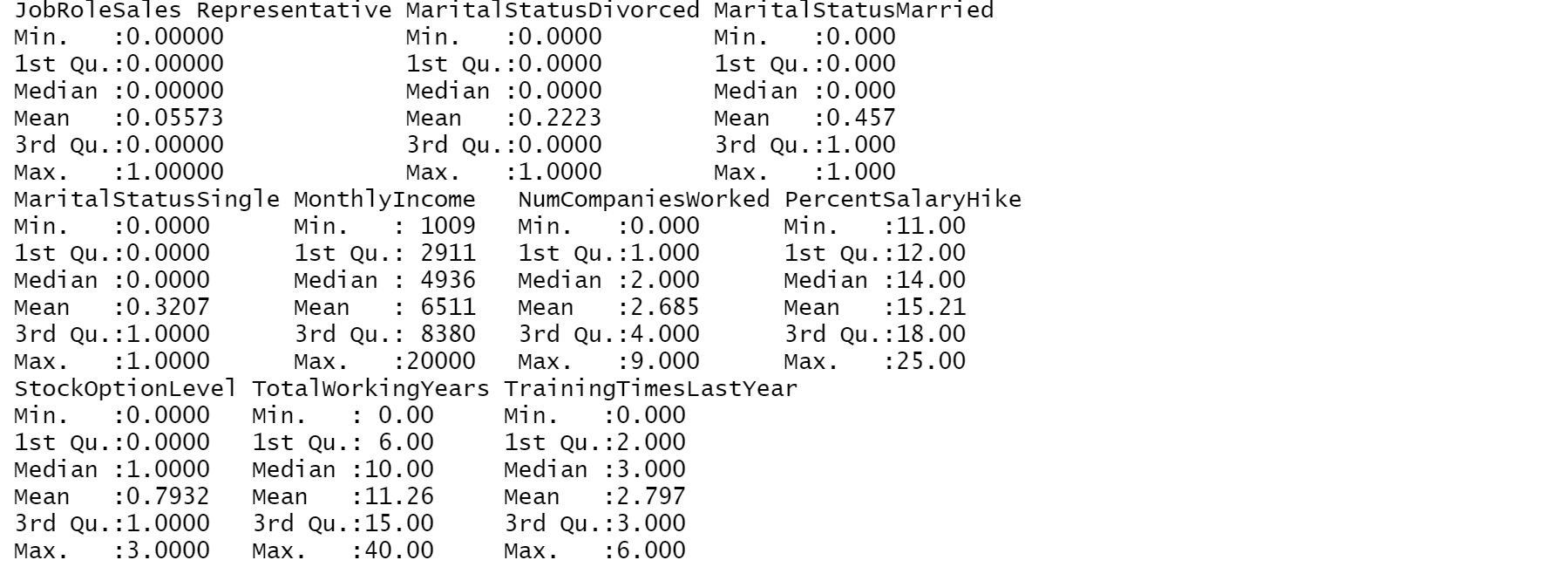
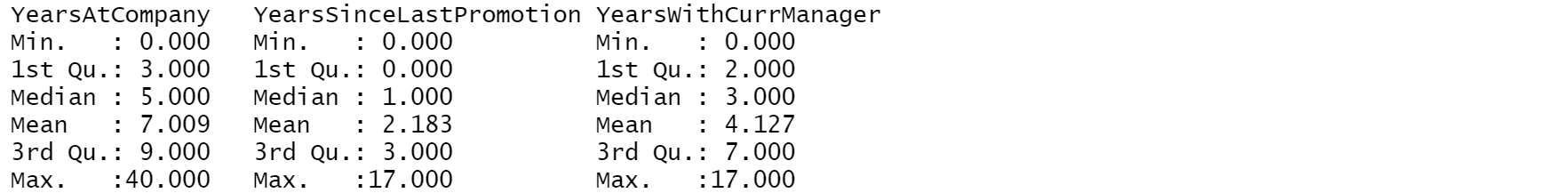
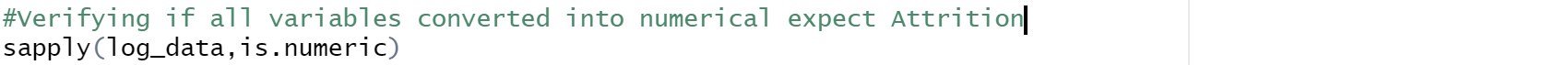
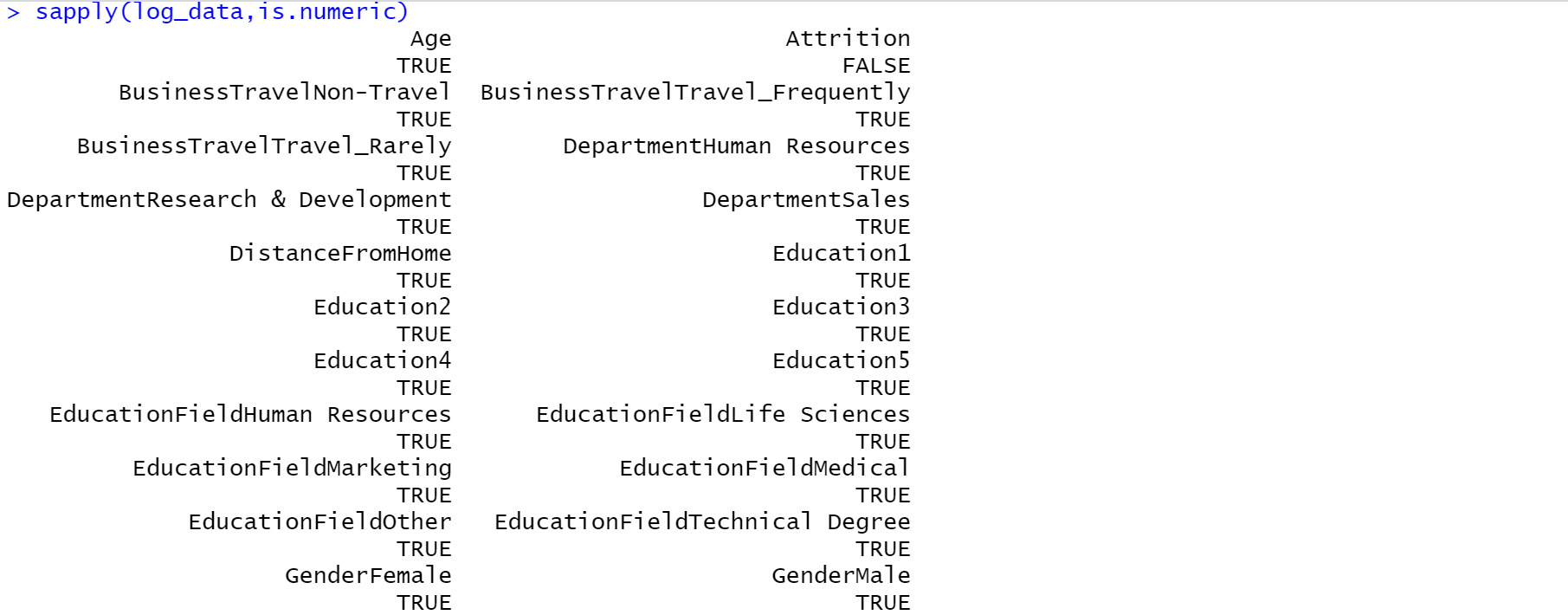
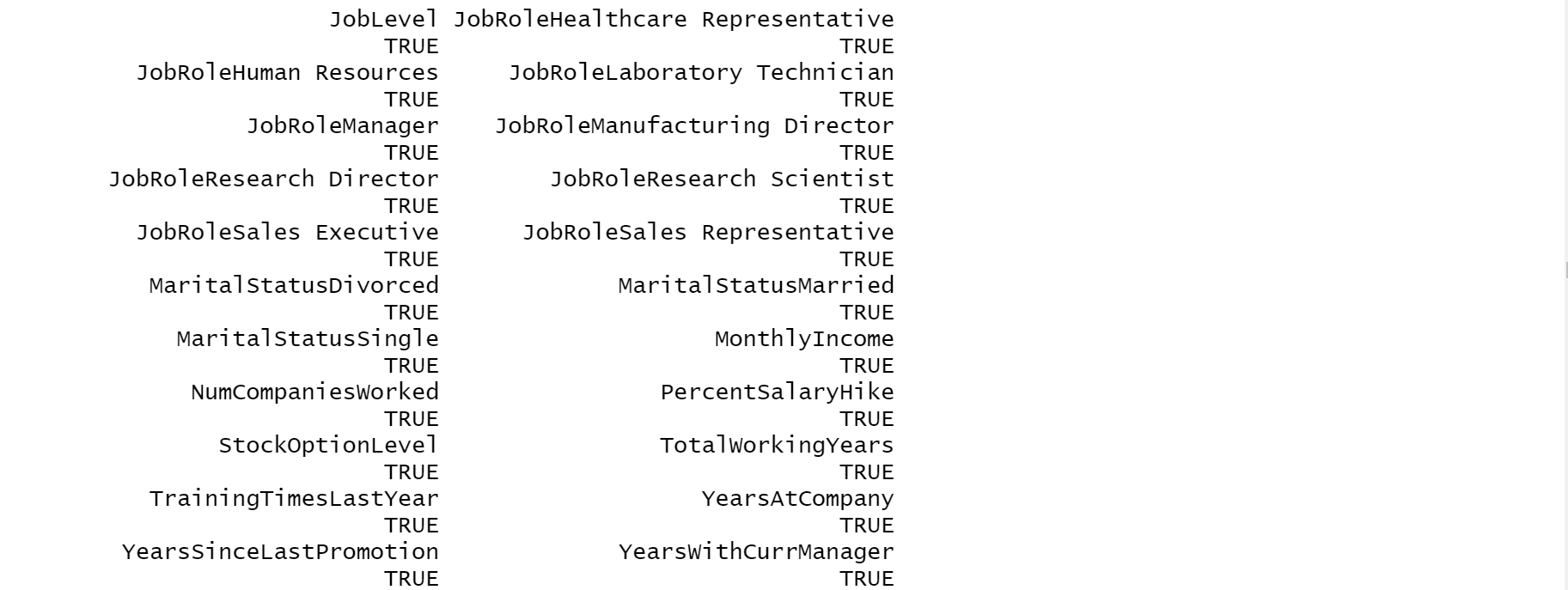
5.3. Building various Classification Models and Evaluating the Best model.

5.3.1: Logistic Regression:

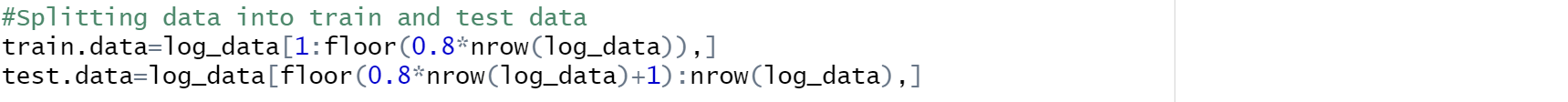
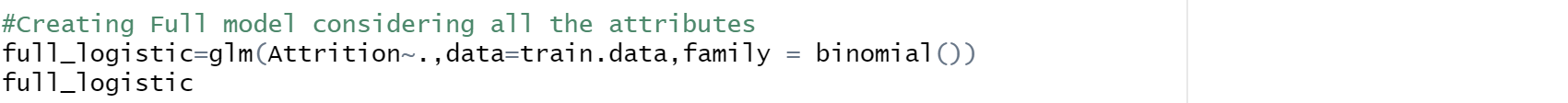
**CREATING DUMMY VARIABLES (TRANSFORMING CATEGORICAL VARIABLES INTO NUMERIC DUMMY VARIABLES)**

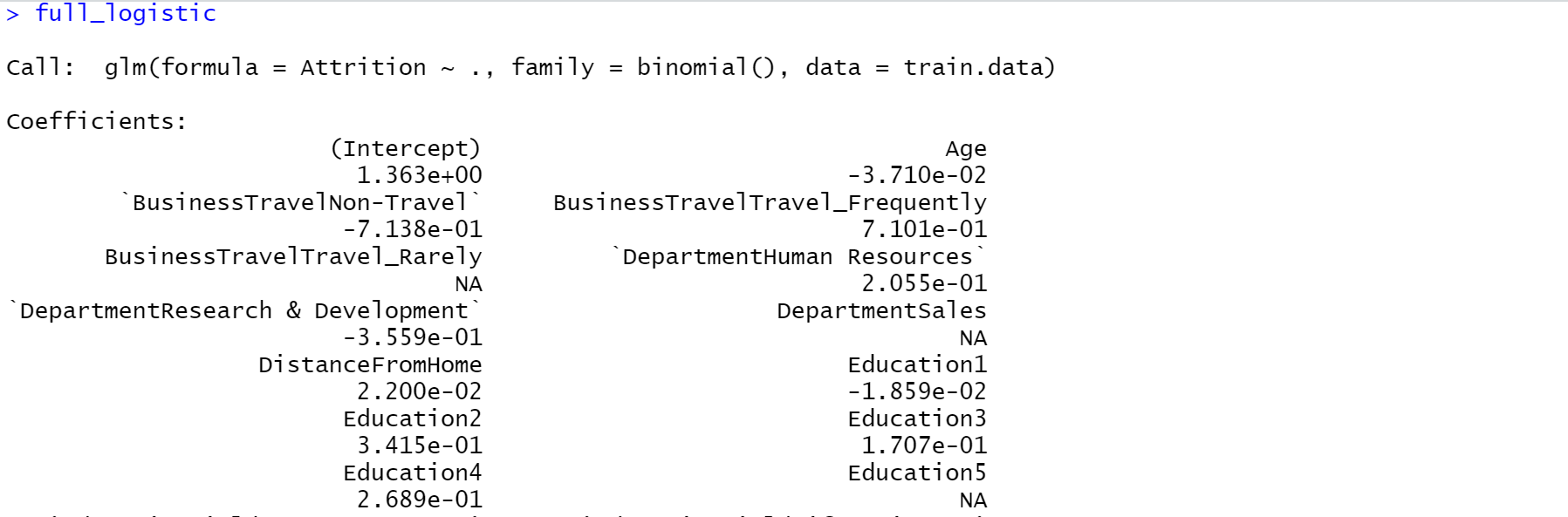
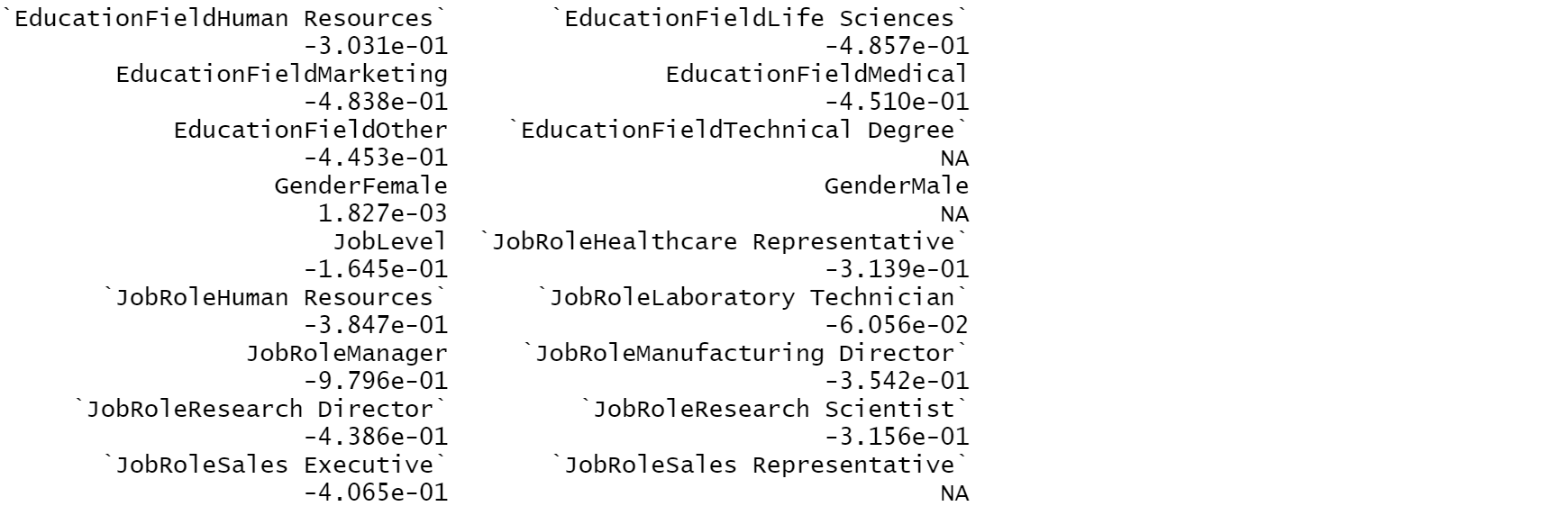
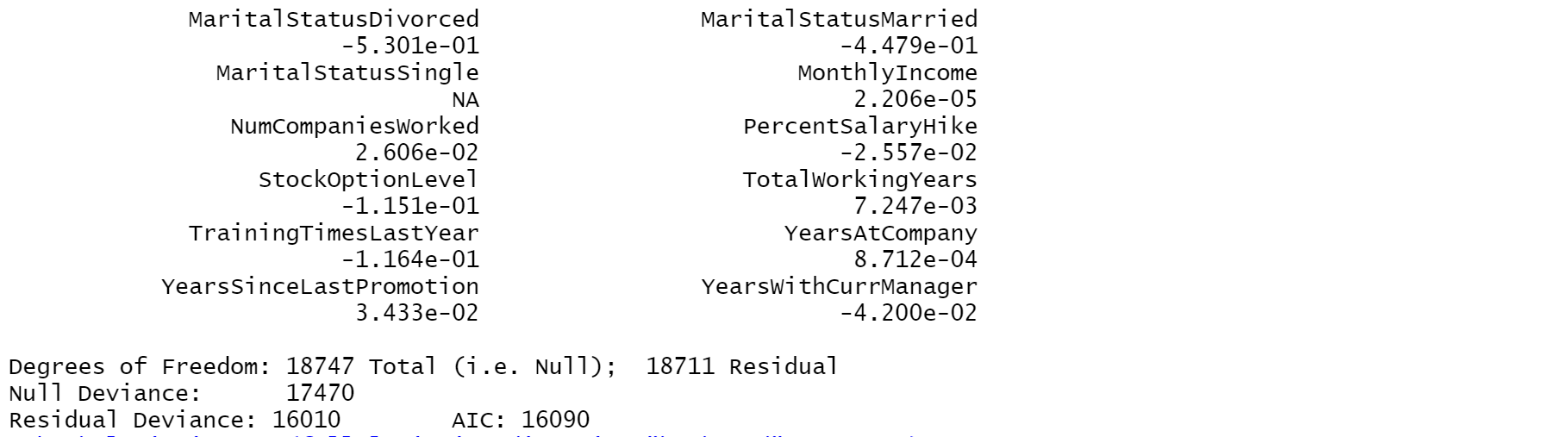
     

**VERIFY IF ALL VARIABLES ARE CONVERTED INTO NUMERICAL EXCEPT ATTRITION:**   

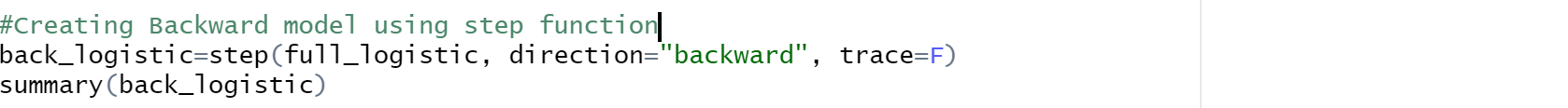
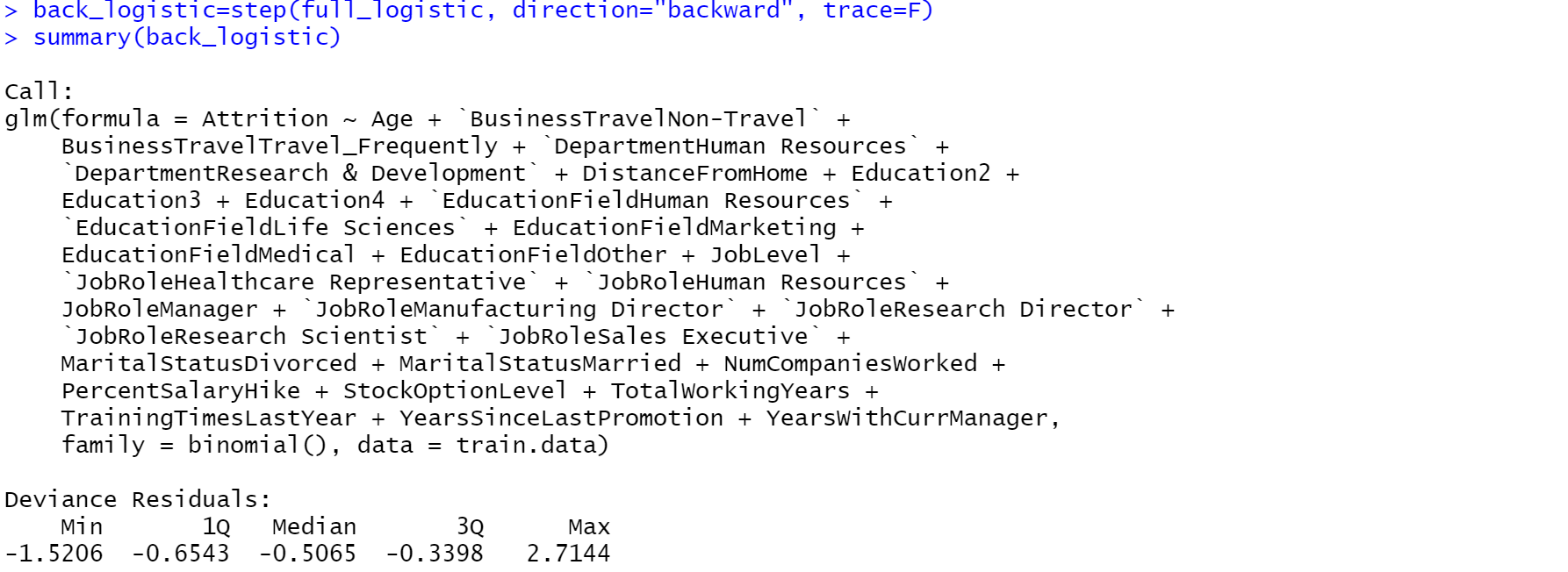
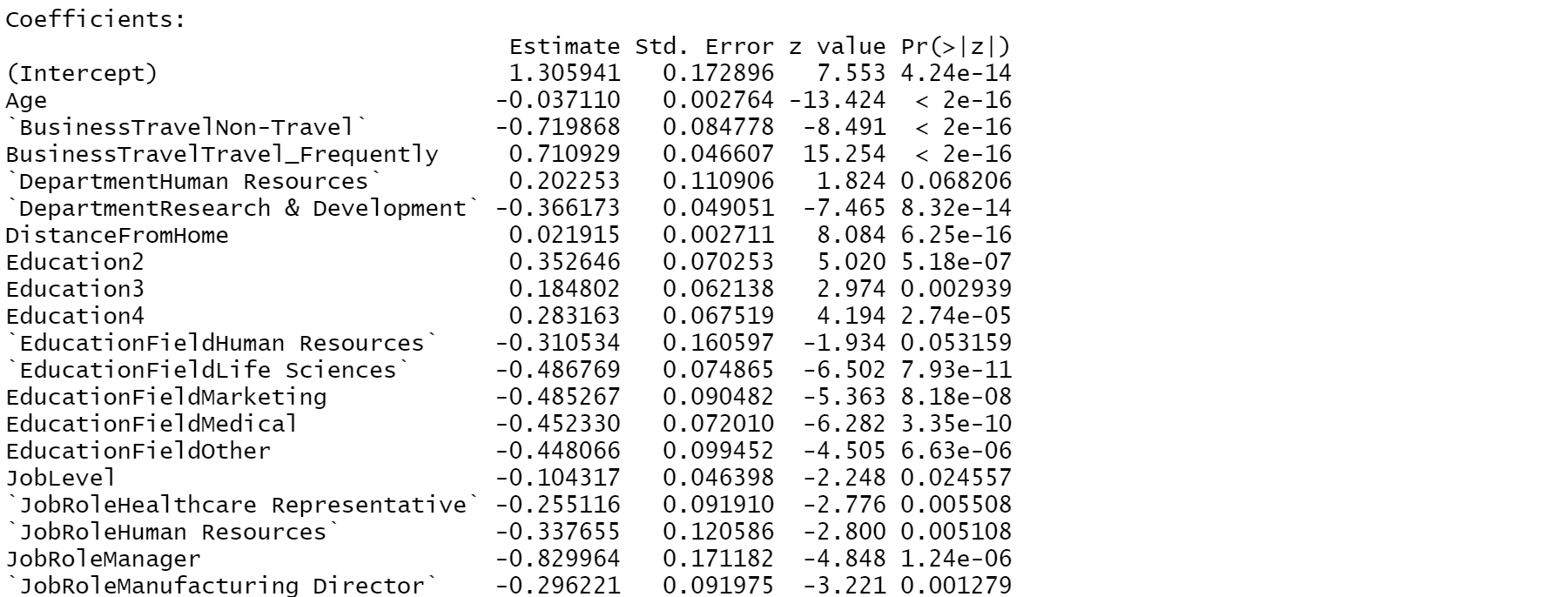
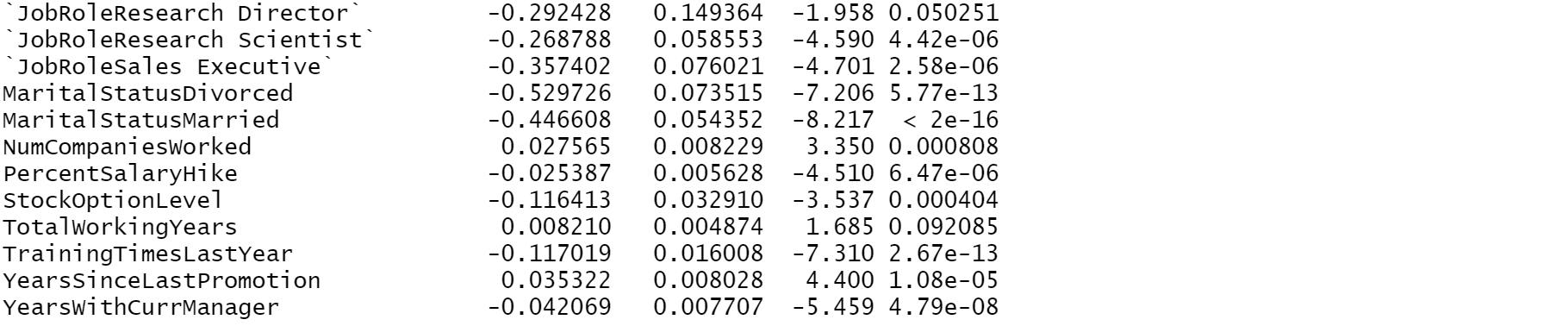
**CREATING FULL MODEL:**

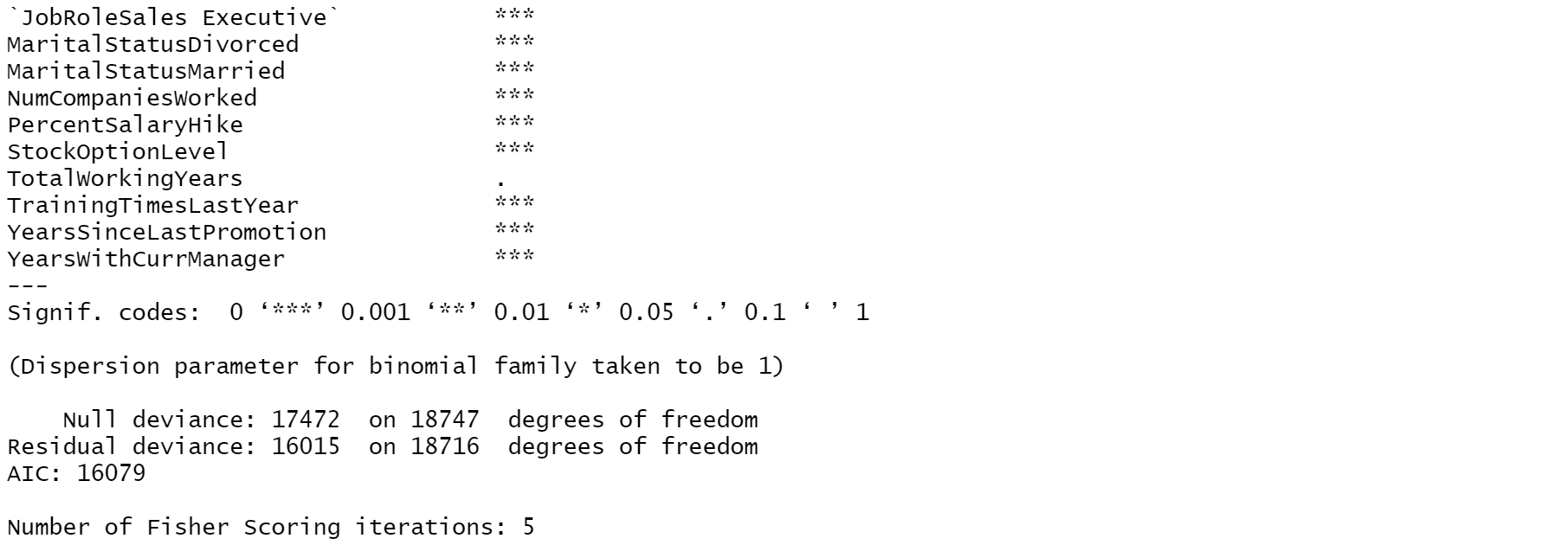
 

**AIC (FULL MODEL): 16090**

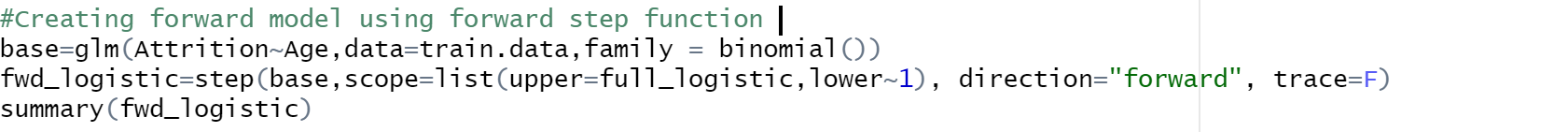
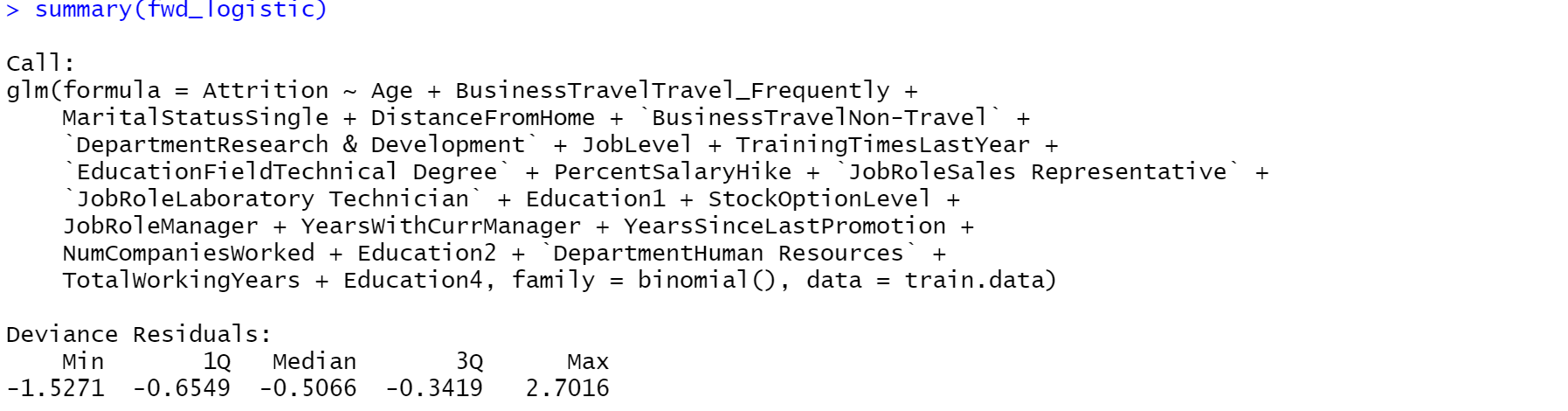
**CREATING BACKWARD MODEL:**

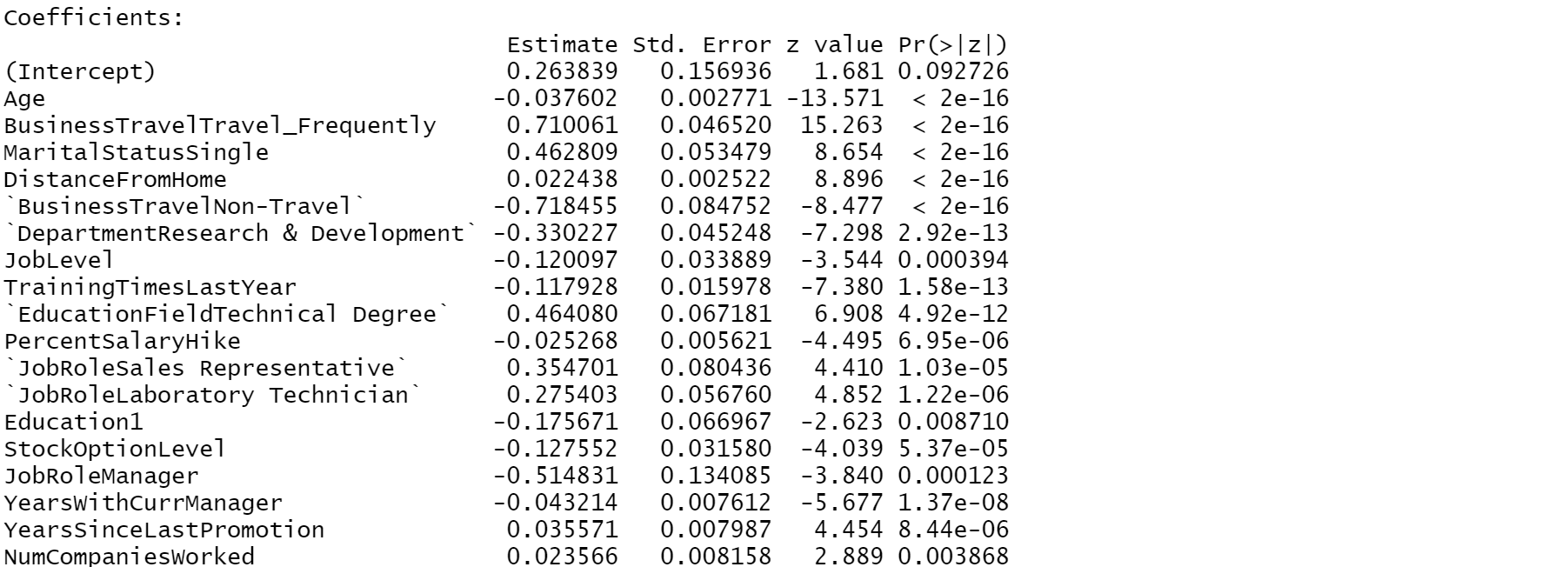
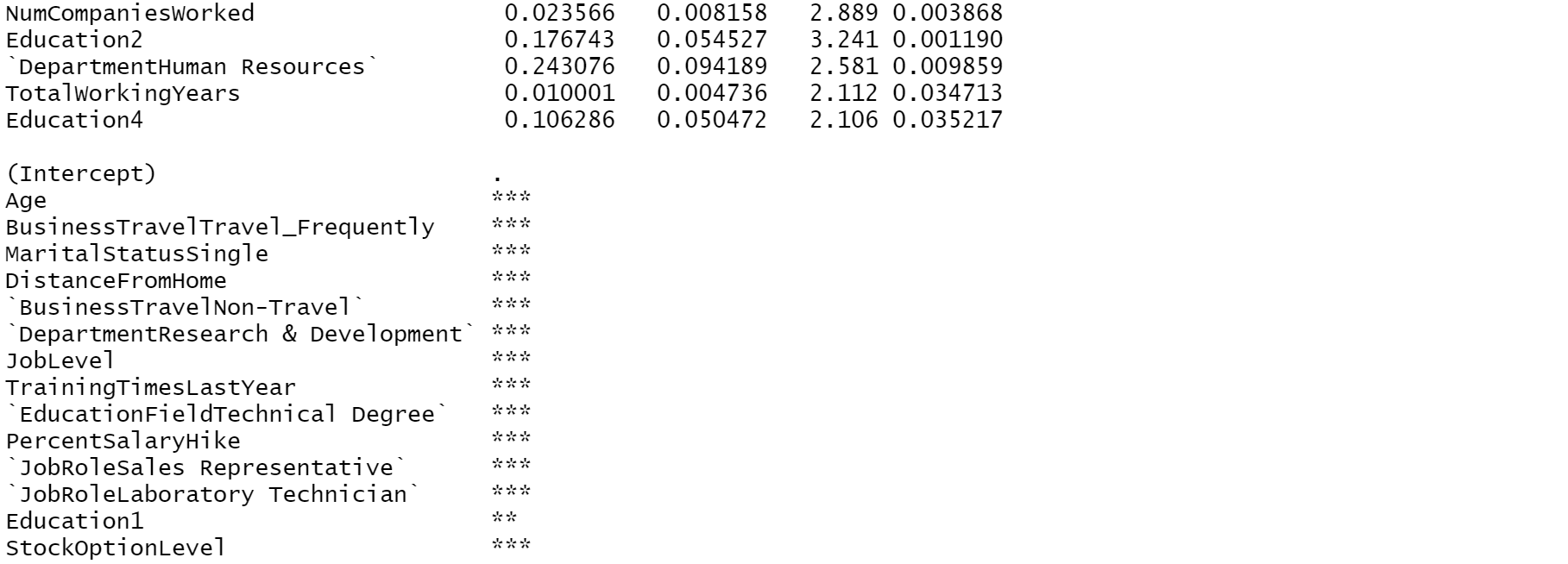
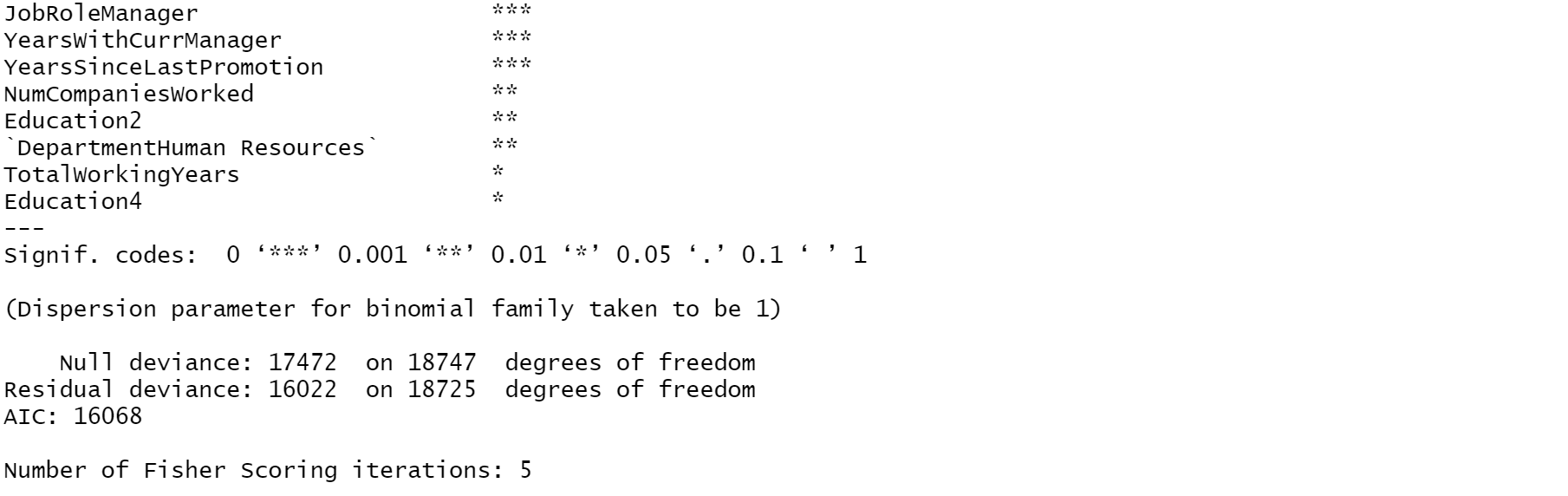
   

**AIC (BACKMODEL): 16079**

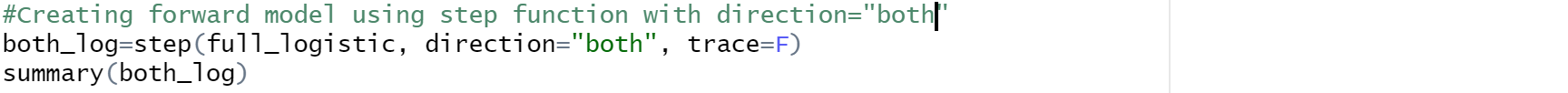
**CREATING FORWARD MODEL**

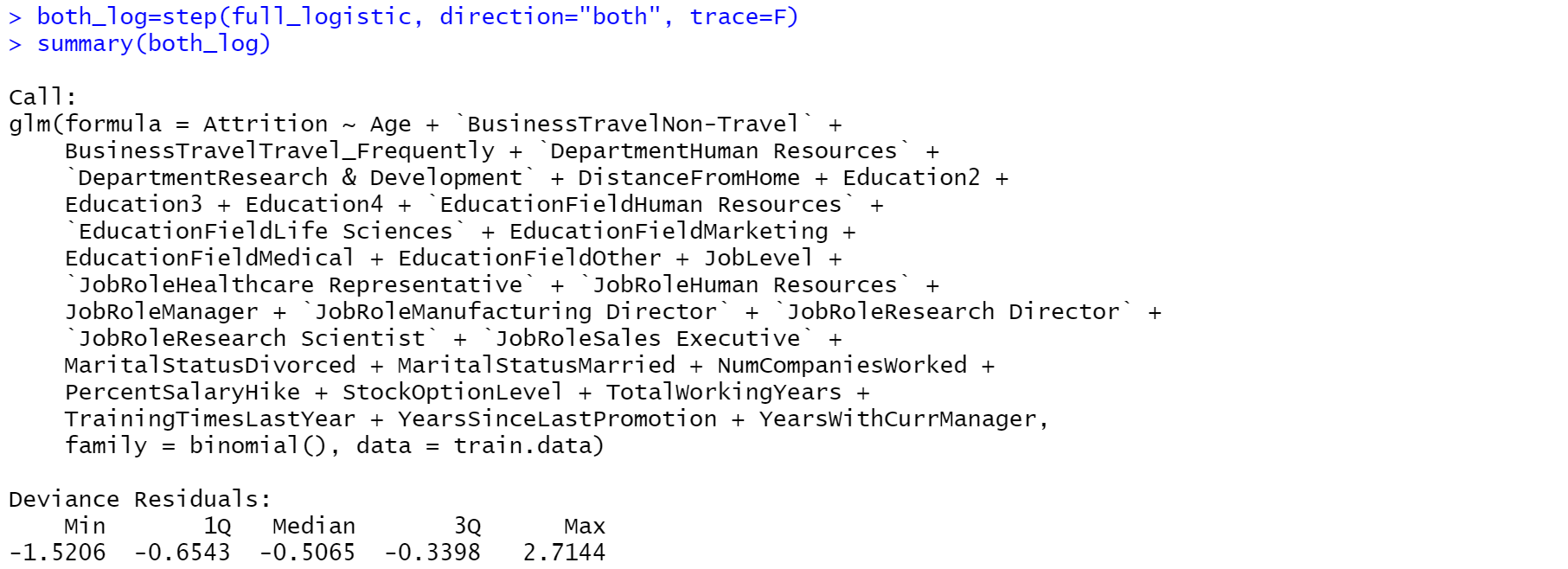
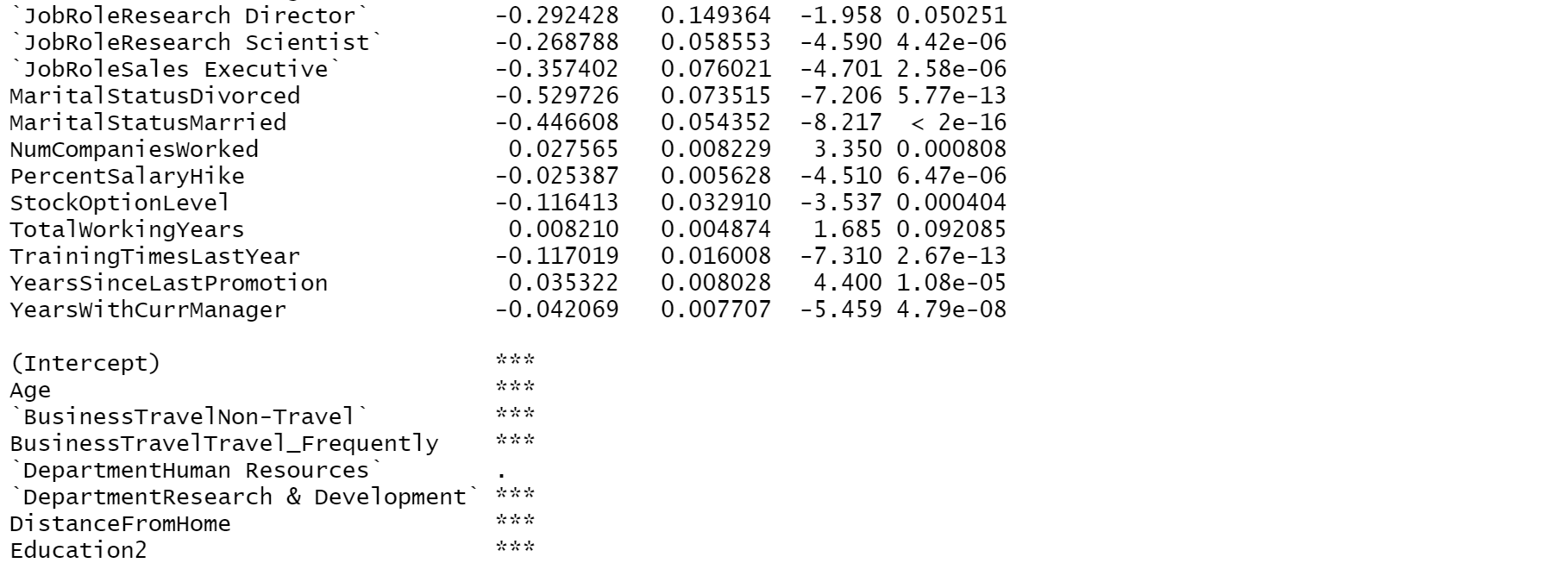
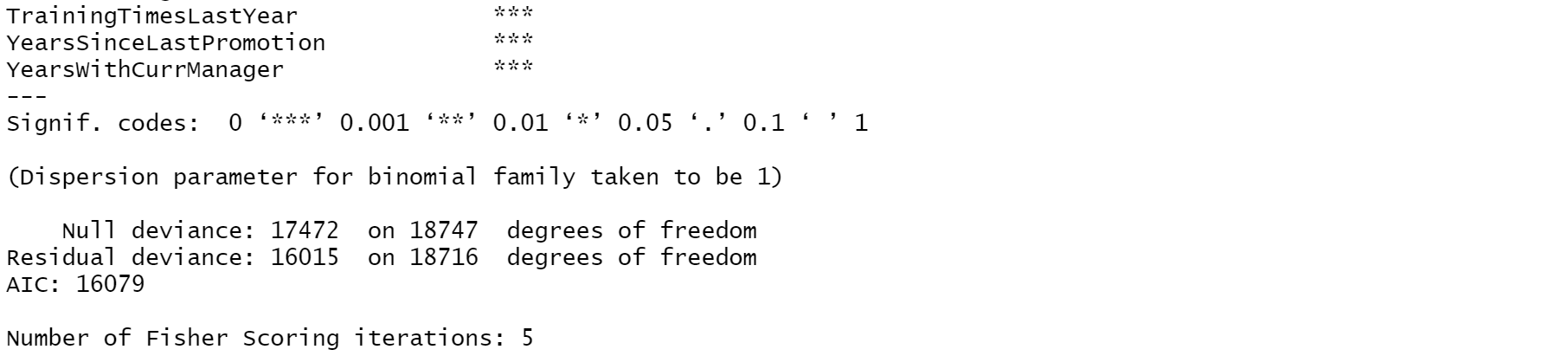
 

**AIC (FORWARD MODEL): 16068**

**CREATING STEPWISE MODEL (DIRECTION= ”BOTH”)**



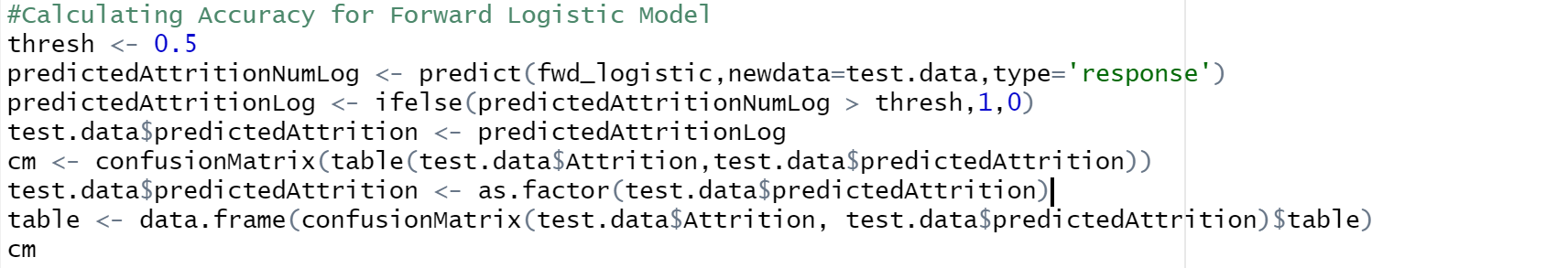
    

**AIC (STEPWISE): 16079**

|  |  |
| --- | --- |
| **LOGISTIC MODEL** | **AIC** |
| FULL | 16090 |
| FORWARD | 16068 |
| BACKWARD | 16079 |
| STEPWISE (BOTH) | 16079 |

**Since the AIC value of Forward logistic model is lesser than the other models. So, we will consider Forward Logistic Model as the best logistic model. Now calculating the accuracy for Forward Logistic Model.**

**CALCULATING ACCURACY FOR FORWARD LOGISTIC MODEL**

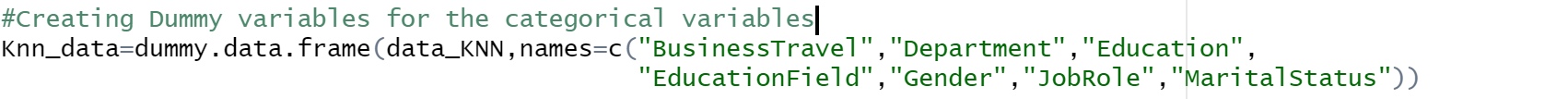
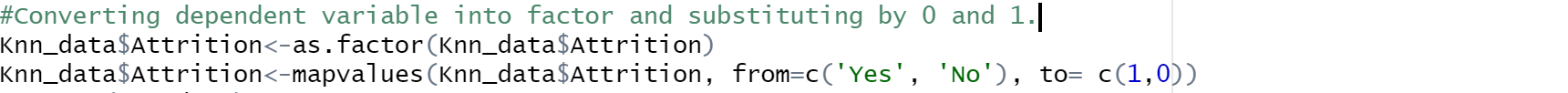
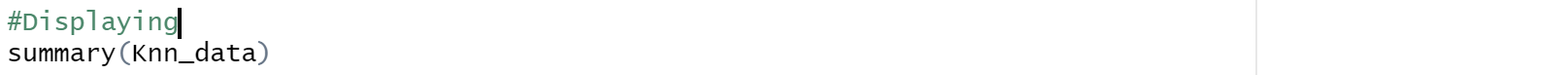
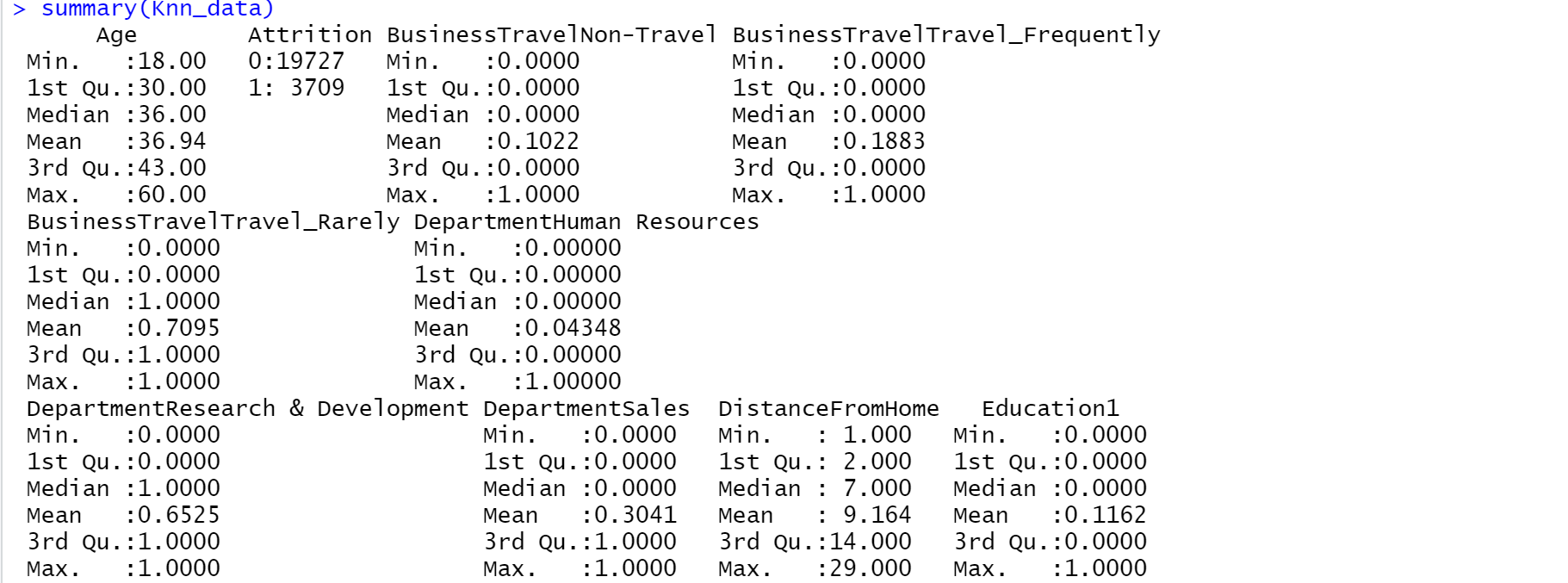
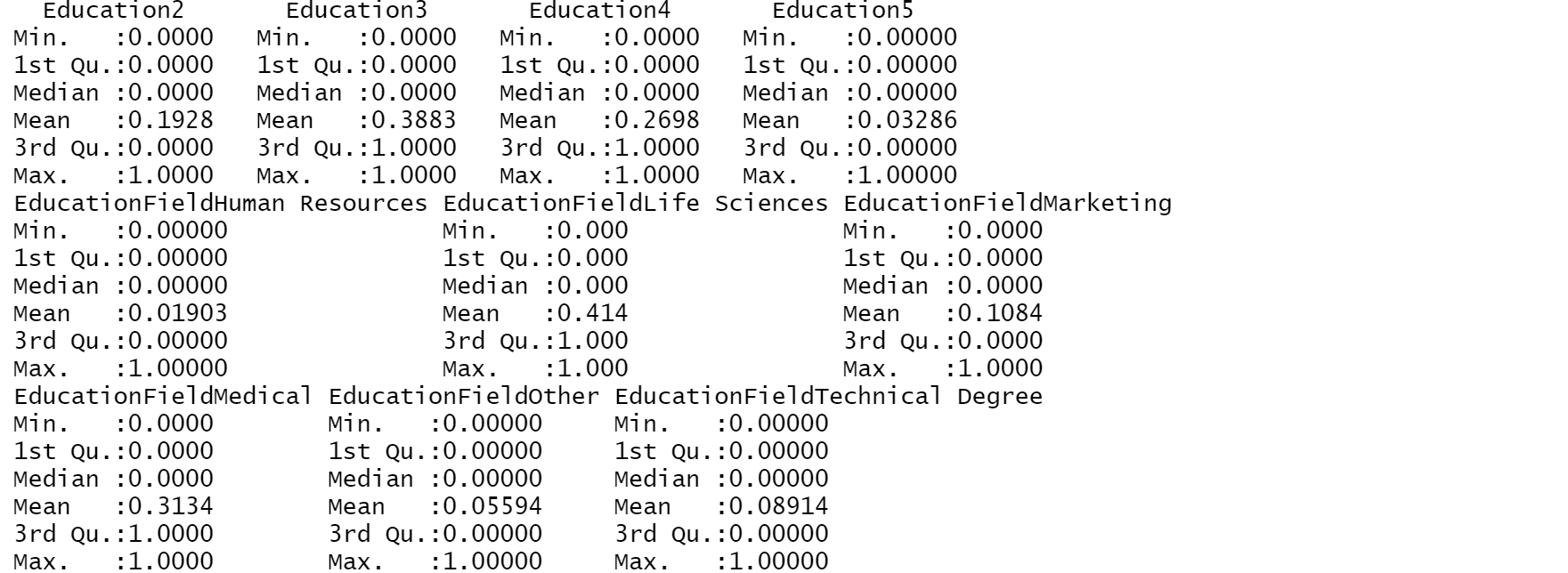
 

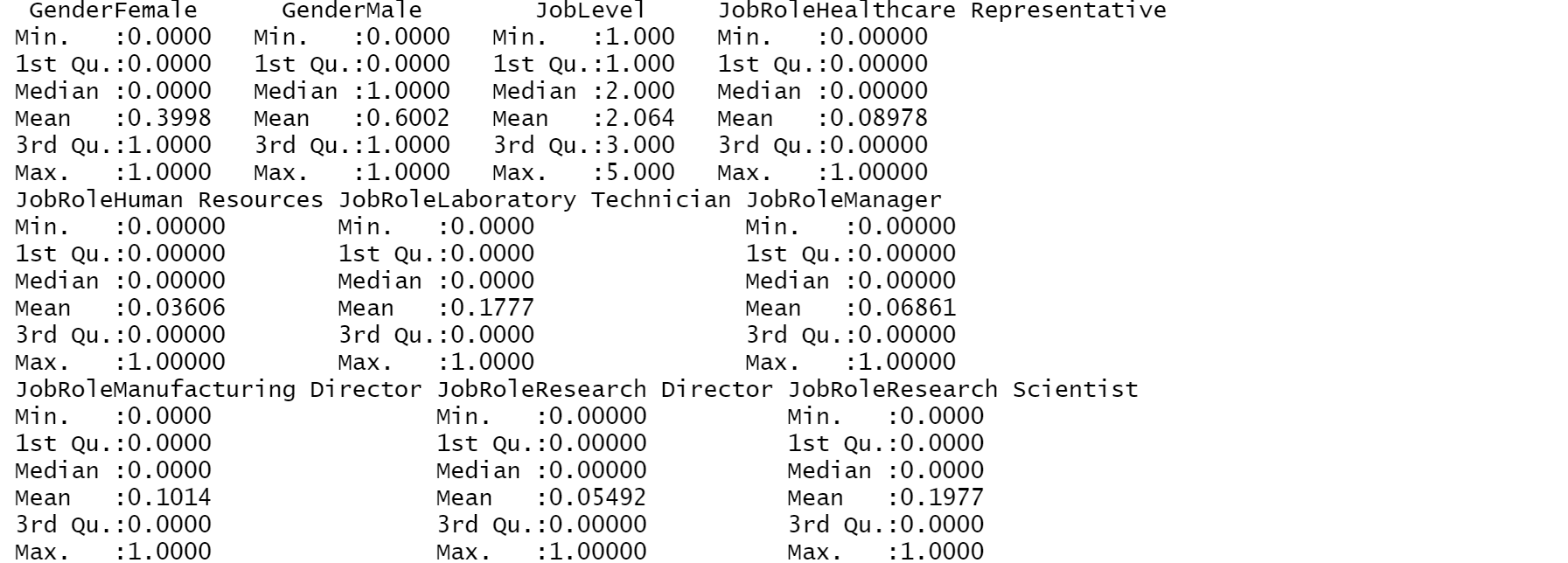
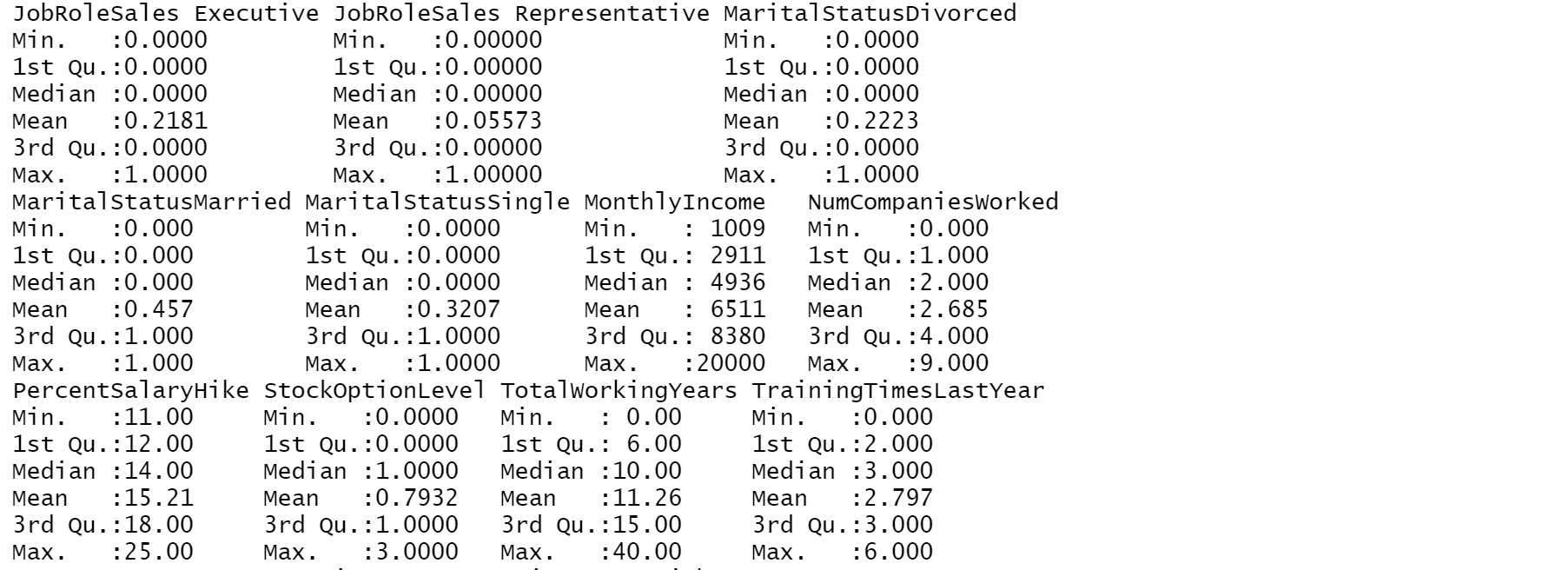
**ACCURACY OF FORWARD LOGISTIC MODEL: 91.19%**

5.3.2 KEY NEAREST NEIGHBOUR CLASSIFICATION MODEL (KNN)

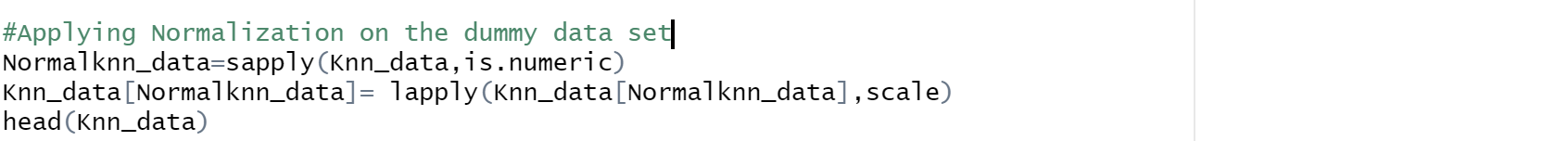
**CREATING DUMMY VARIABLES:**

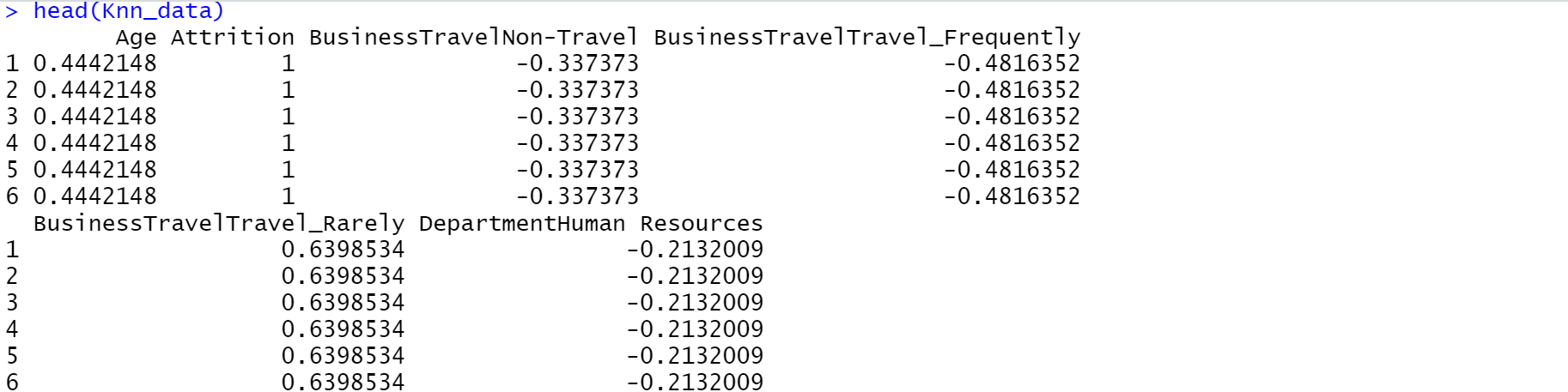


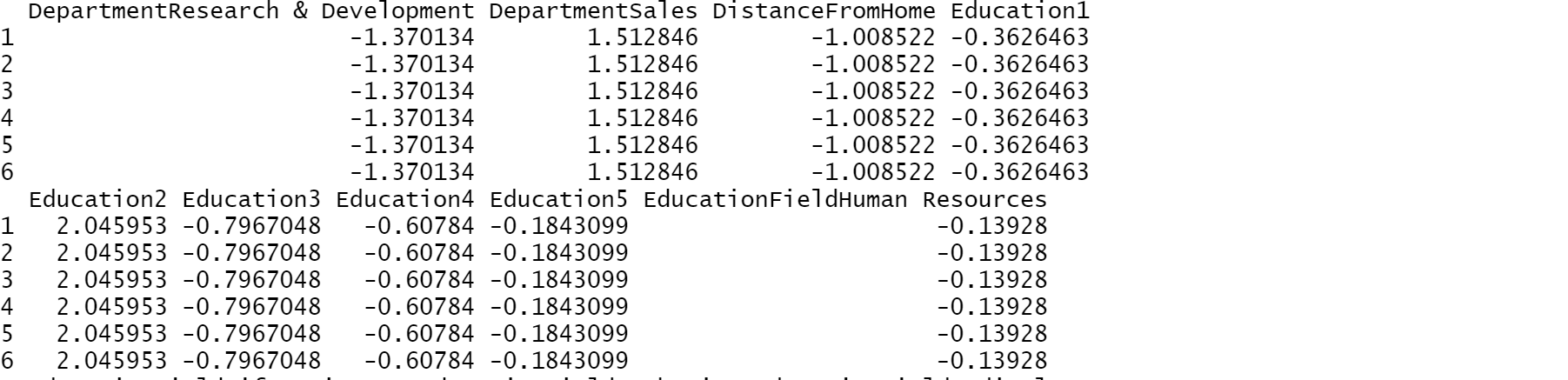
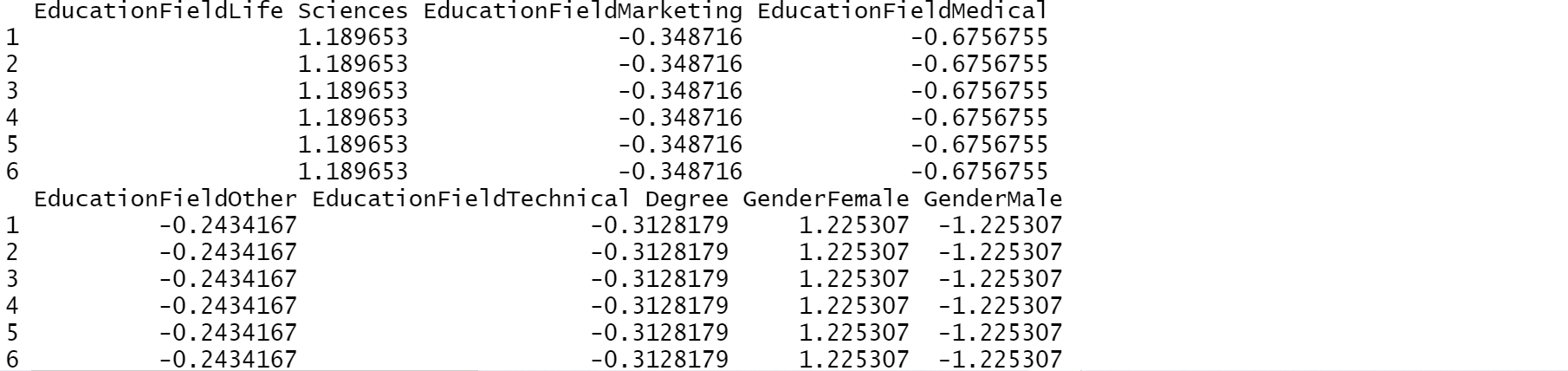
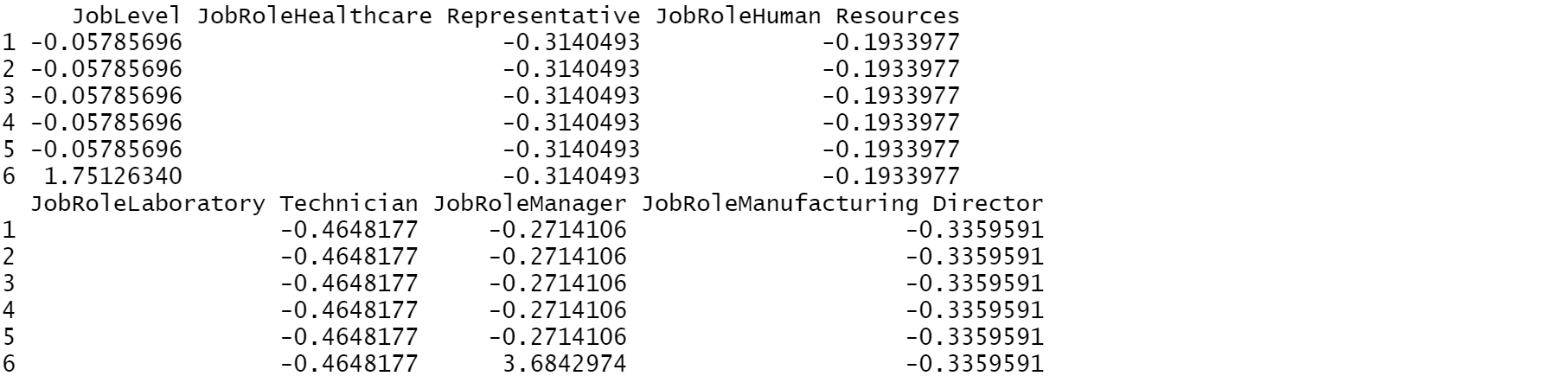
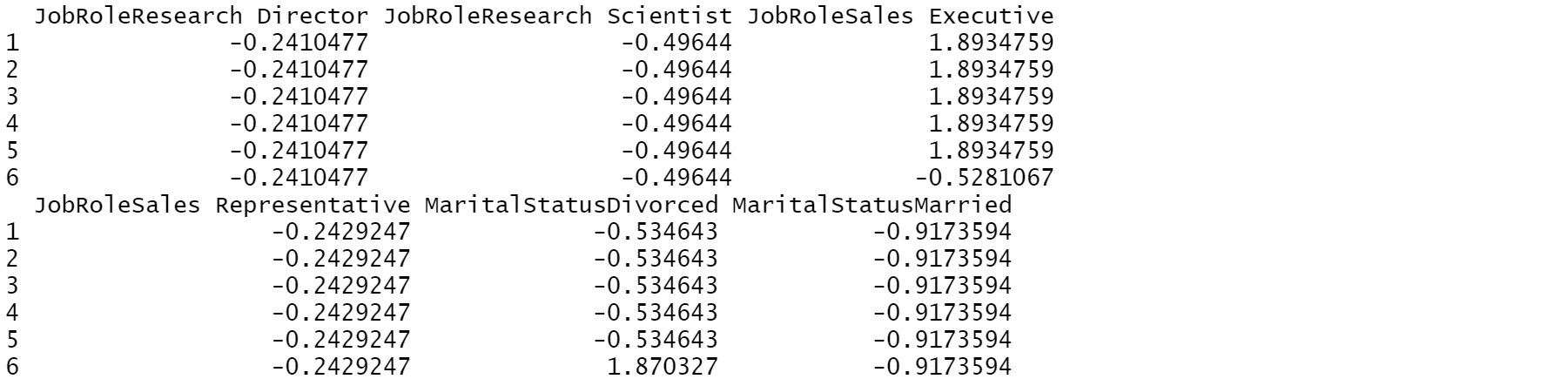
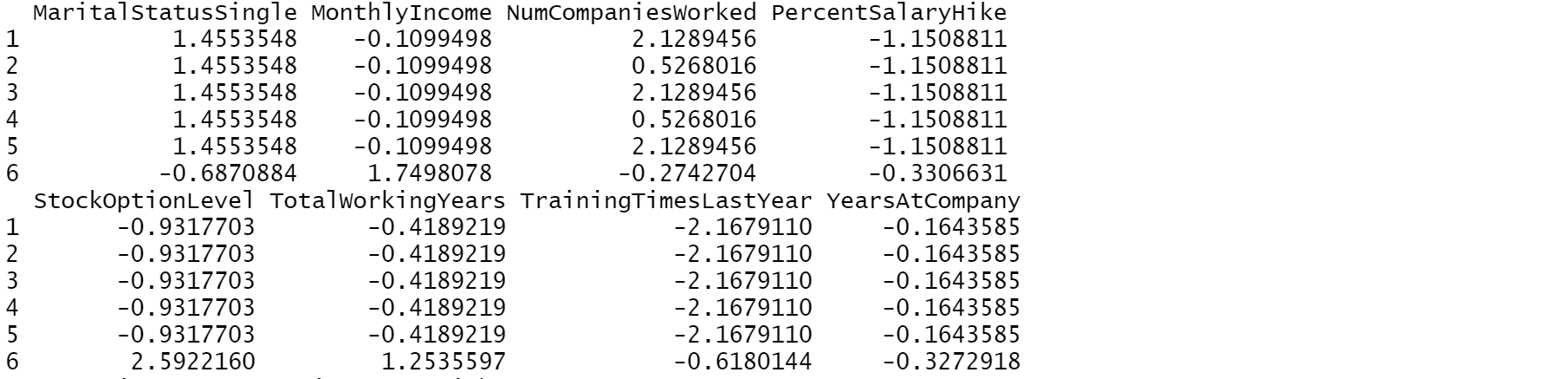
    

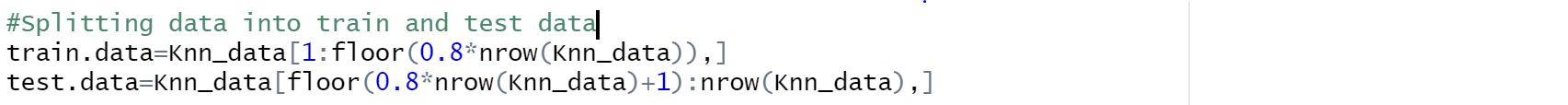
  

**NORMALIZING THE DATA:**

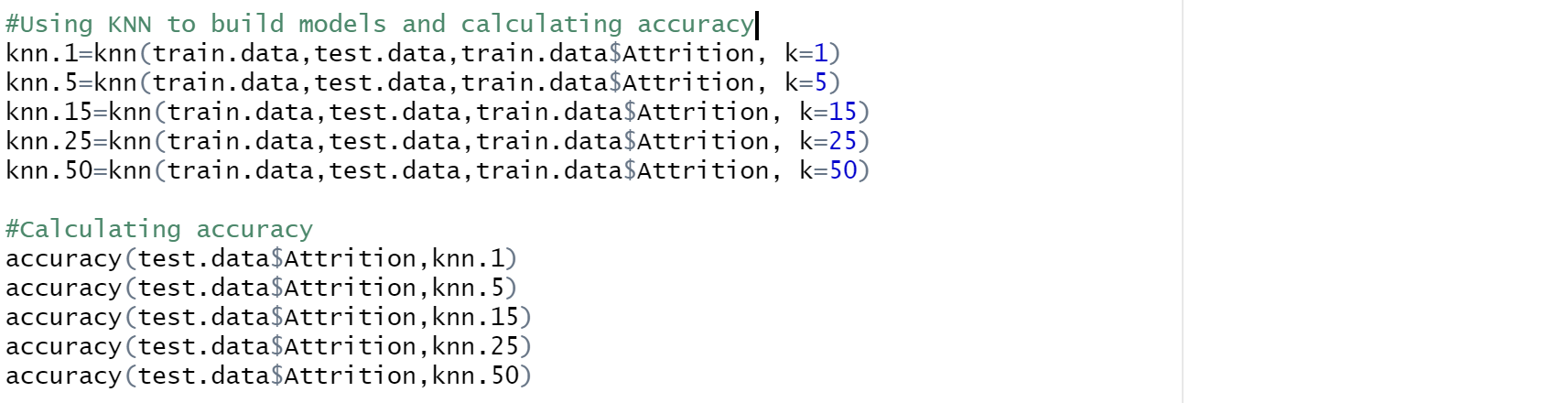






**BUILDING KNN MODELS AND CALCULATING ACCURACY USING HOLD OUT EVALUATIONS**

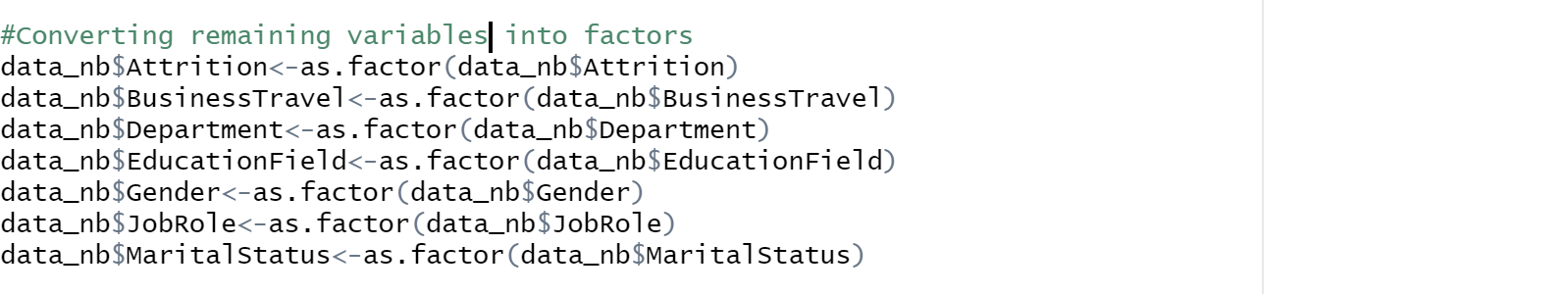
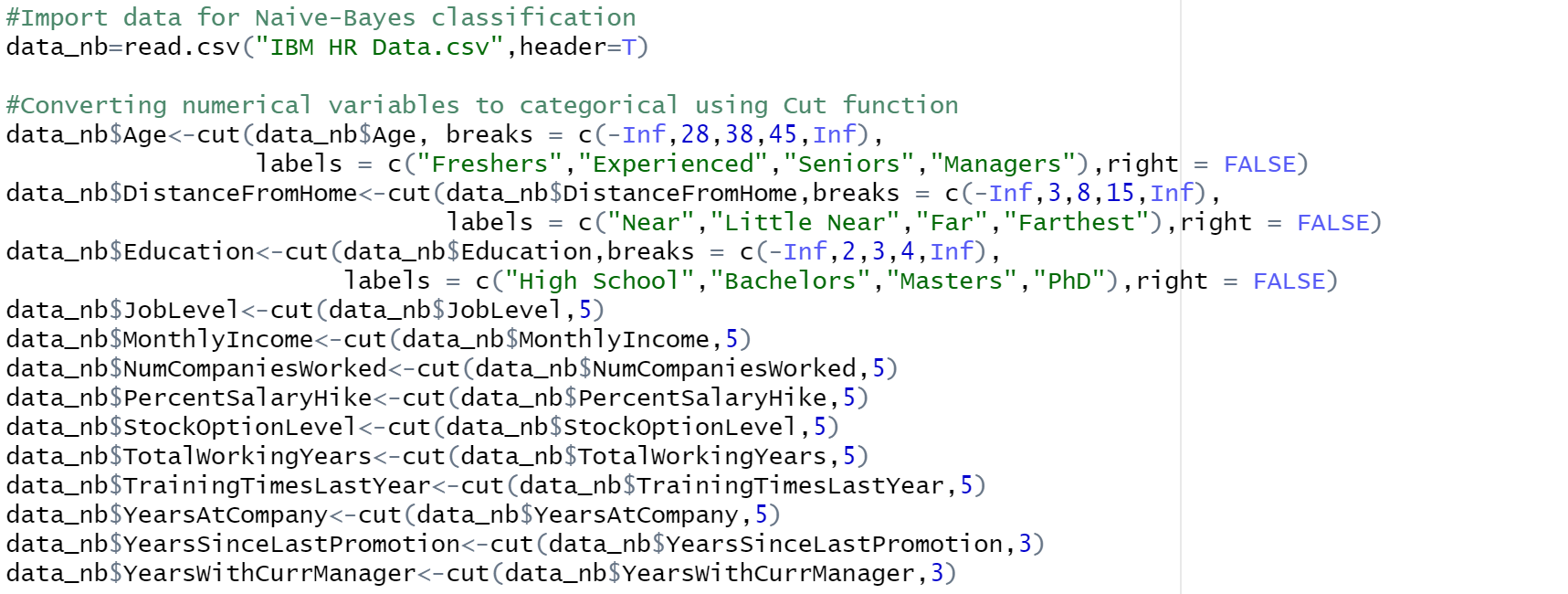
 

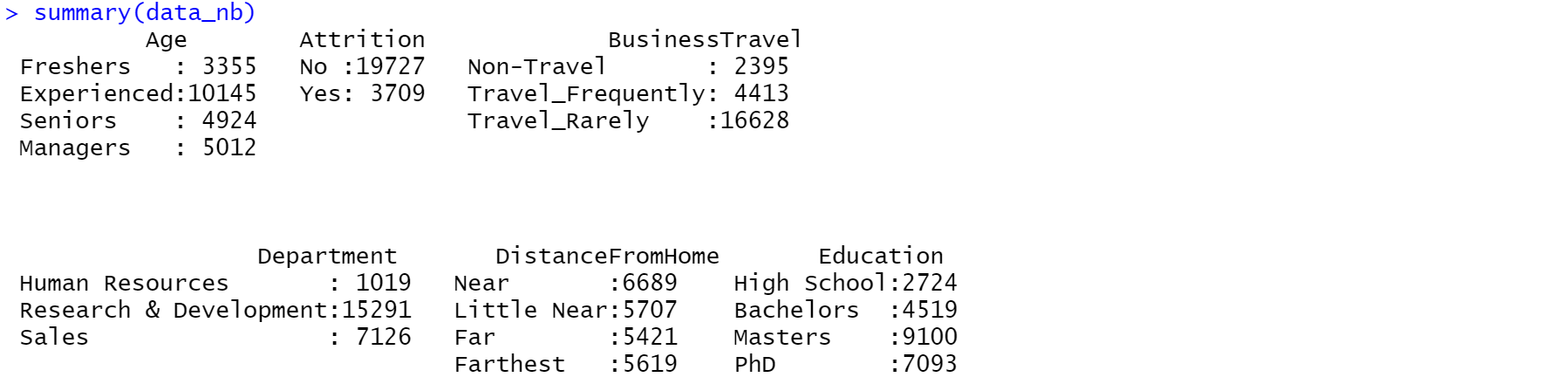
As the K value increases the accuracy value is increasing. But as K value increases rate of error also increases. With K=1 the model achieved 83.85% accuracy with least error rate. When K=5 the accuracy decreases and after that increasing K value there is not much improvement in the accuracy percentage. Therefore, we consider best KNN model with least error percentage i.e. when K=1 and accuracy = 83.85%.

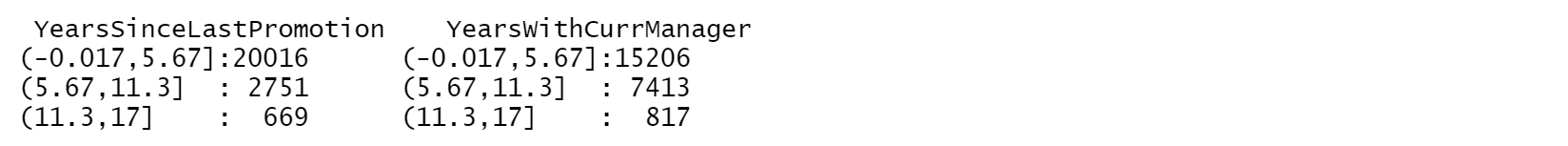
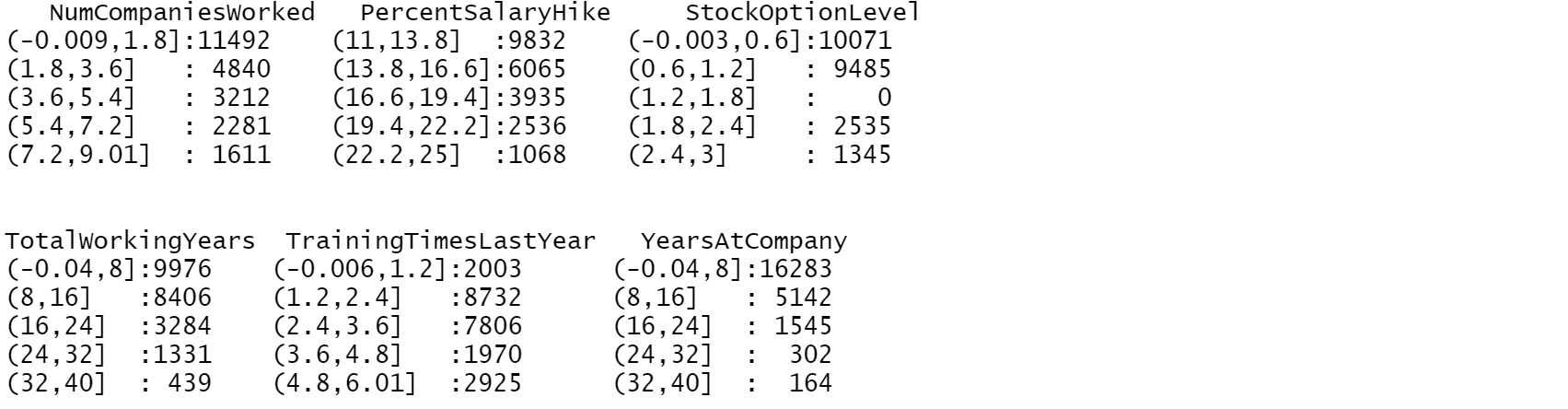
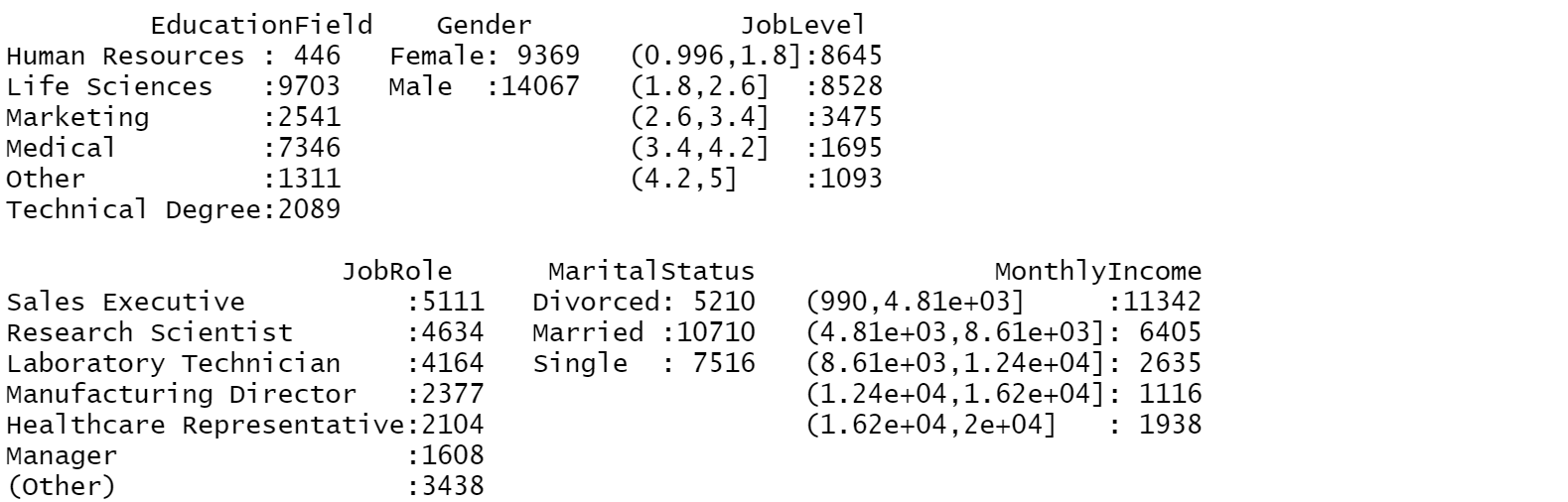
**Accuracy of the model is 83.85%.**

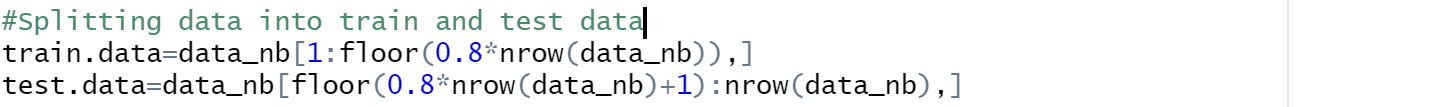
5.3.3 NAÏVE BAYES CLASSIFICATION MODEL

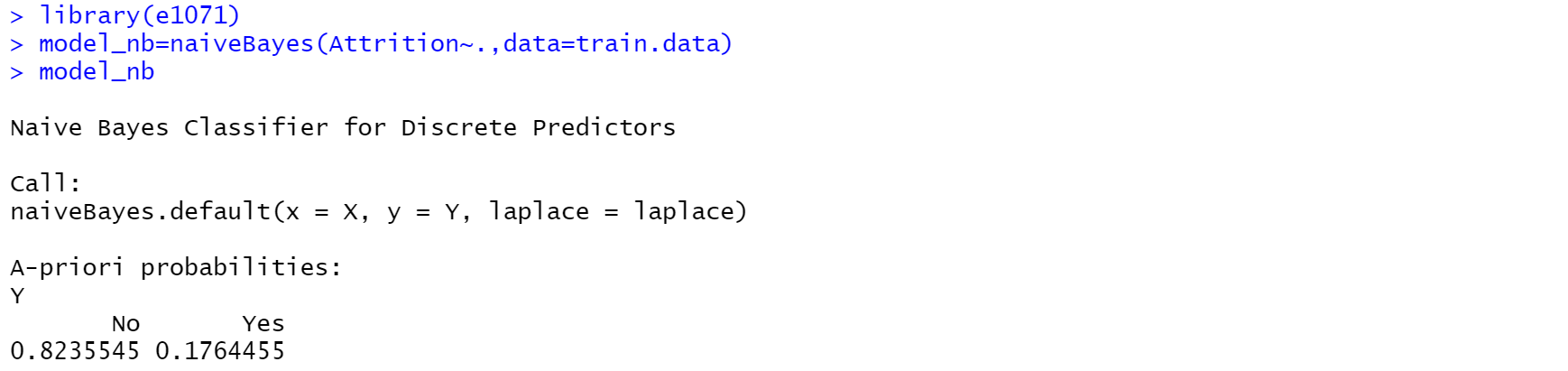
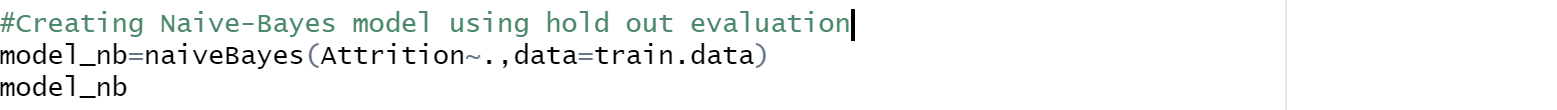
IMPORTING CONVERTING NUMERICAL VARIABLES INTO CATEGORICAL VARIABLES USING CUT FUNCTION

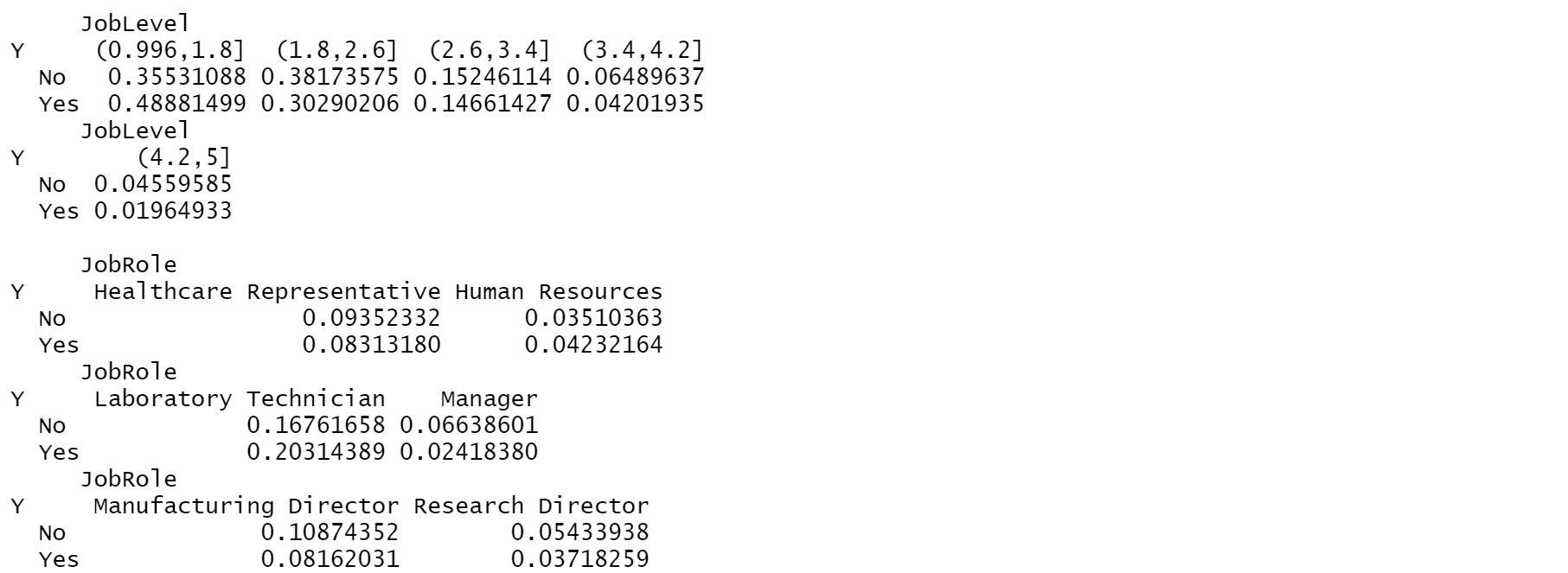
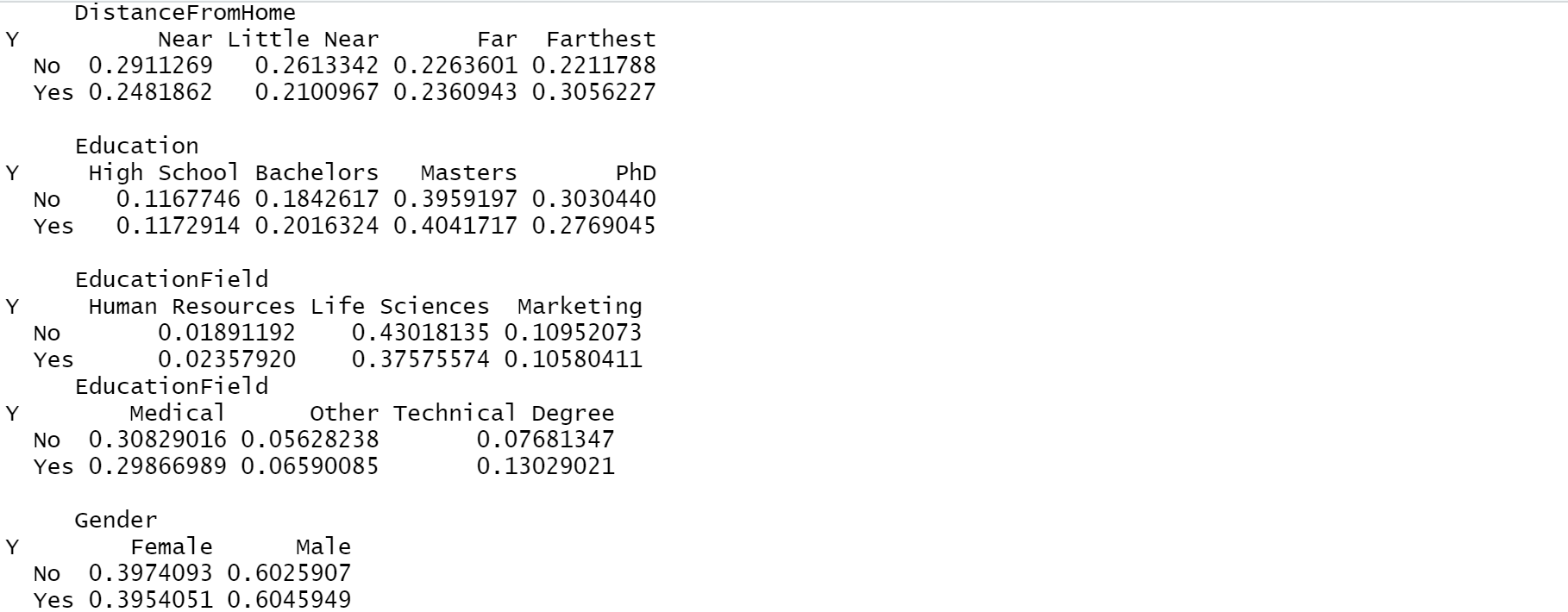
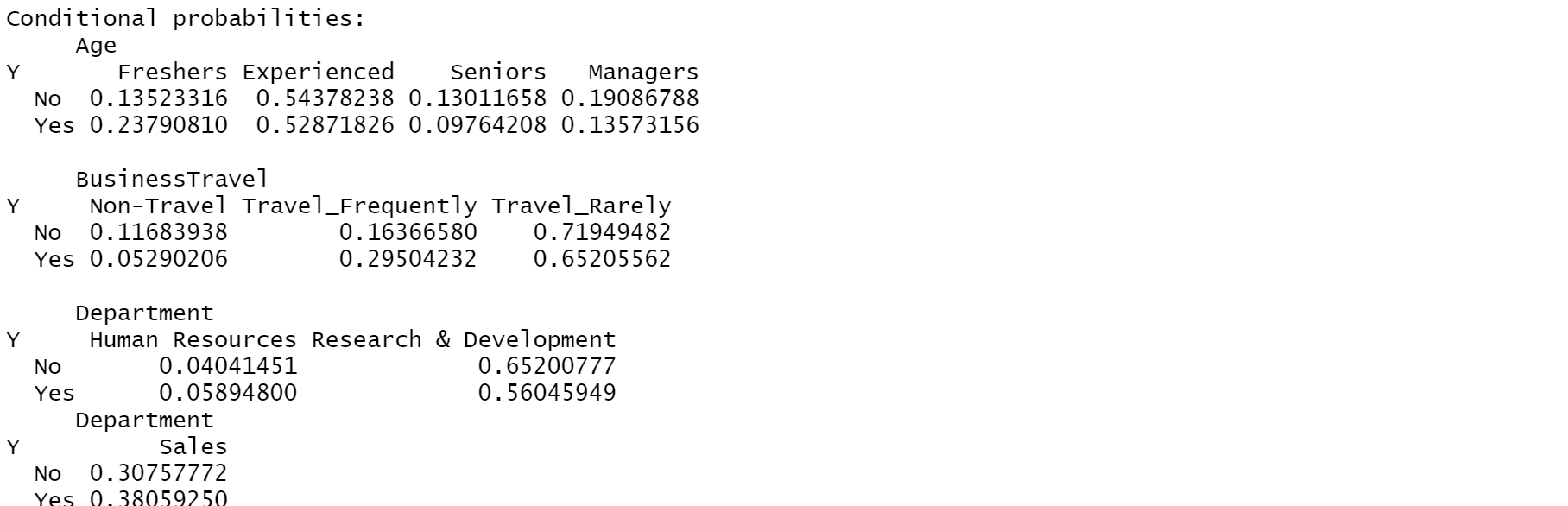


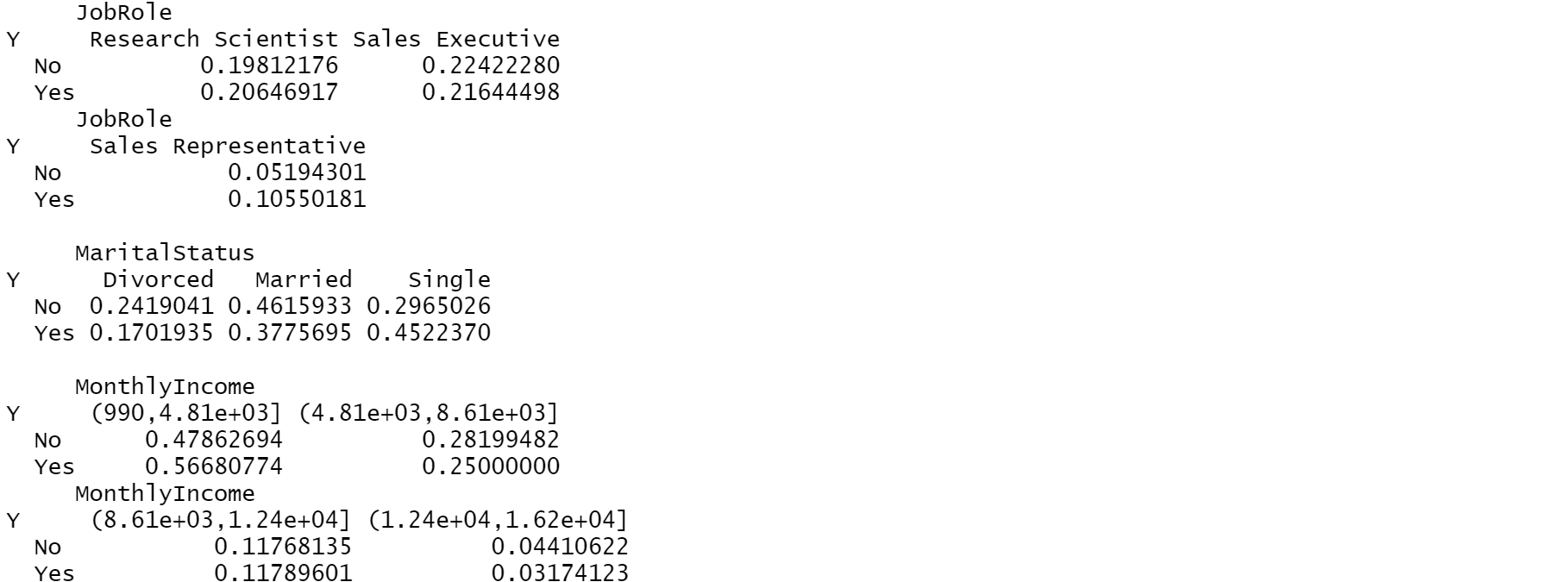


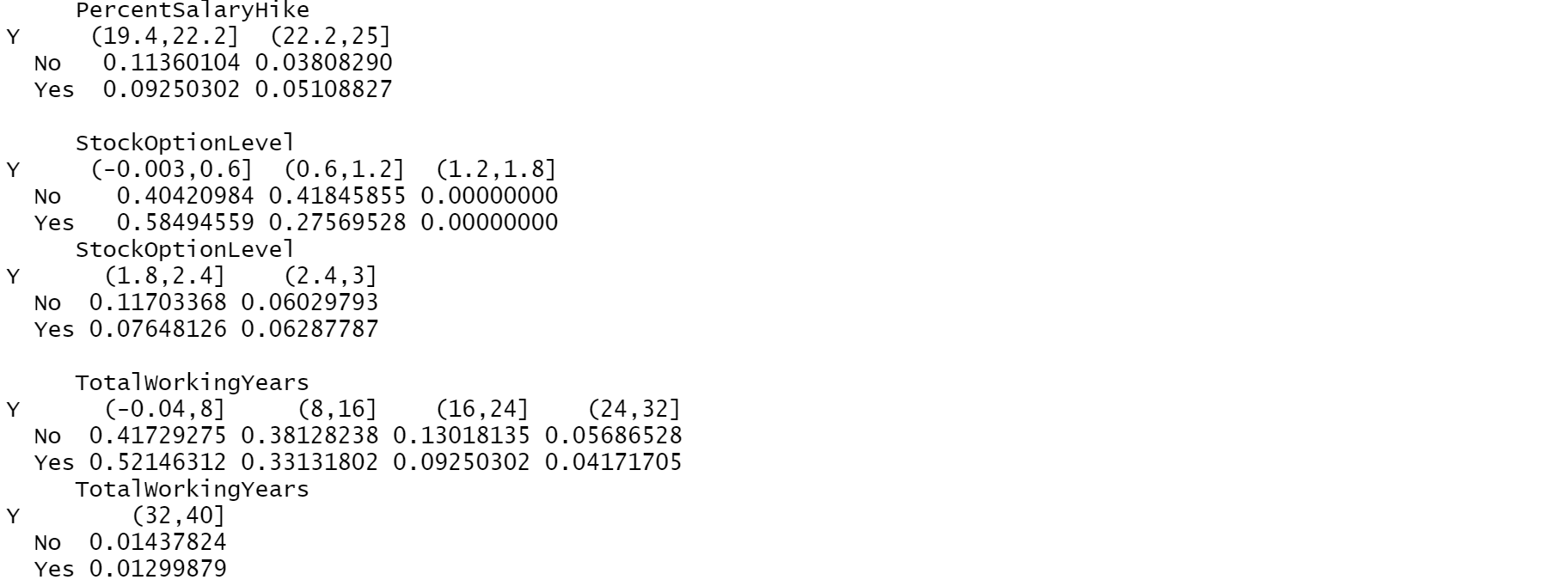
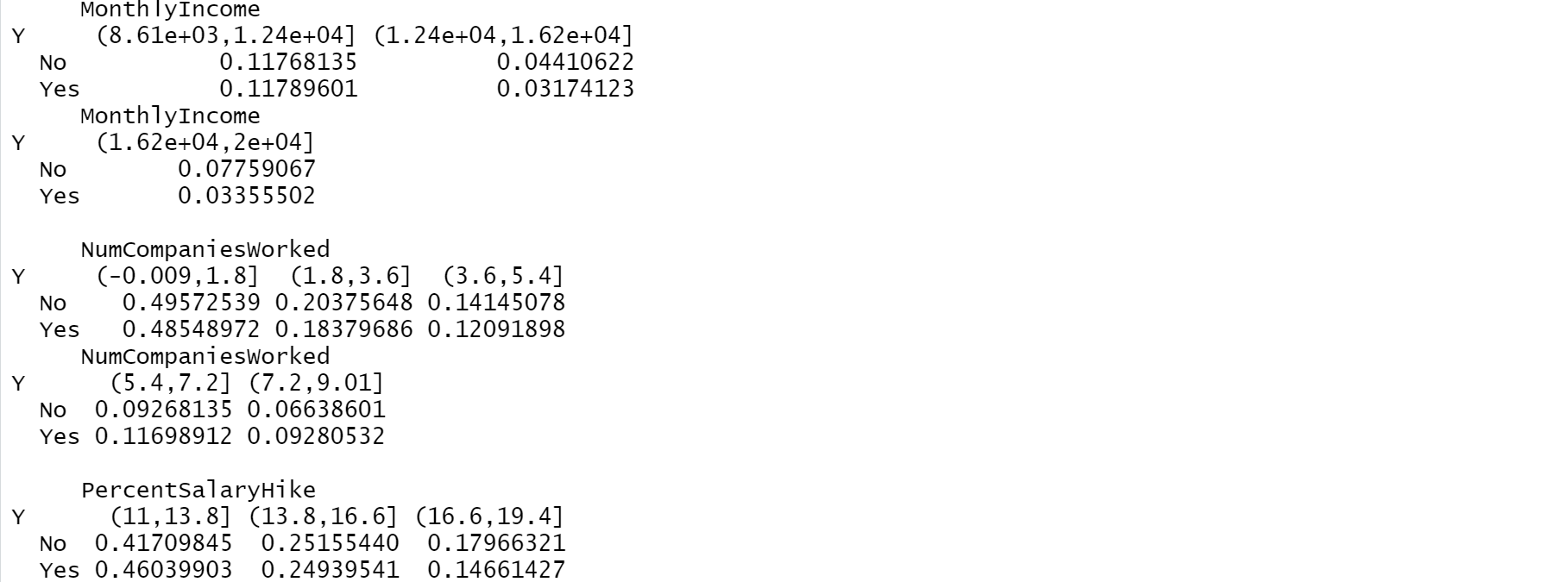




CREATING NAÏVE BAYES MODEL USING HOLD OUT EVALUATION

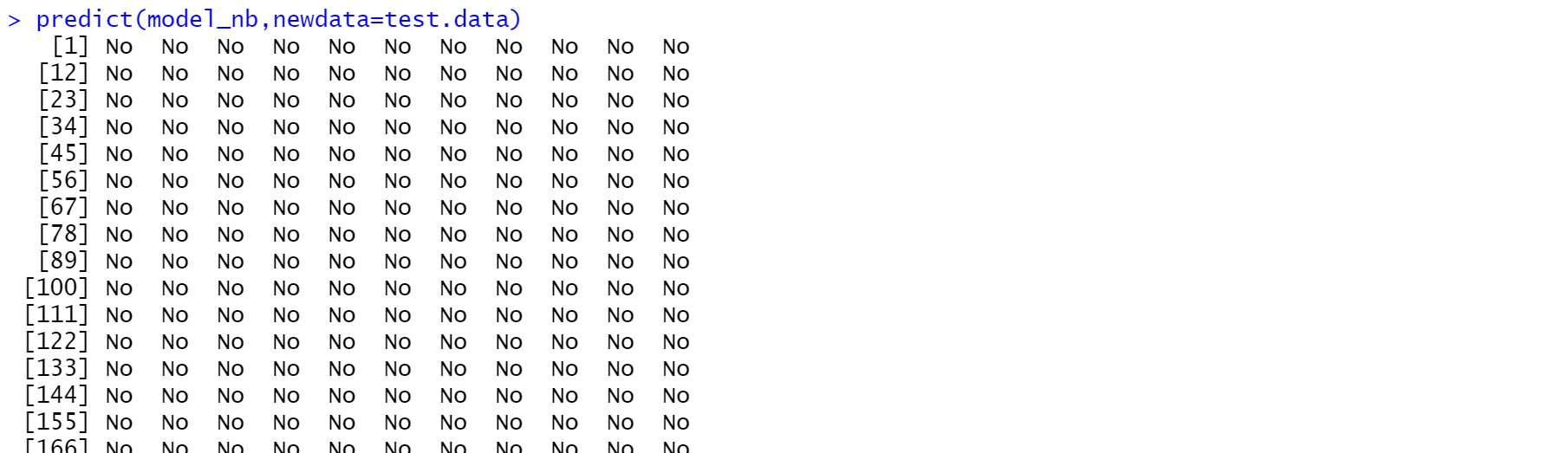
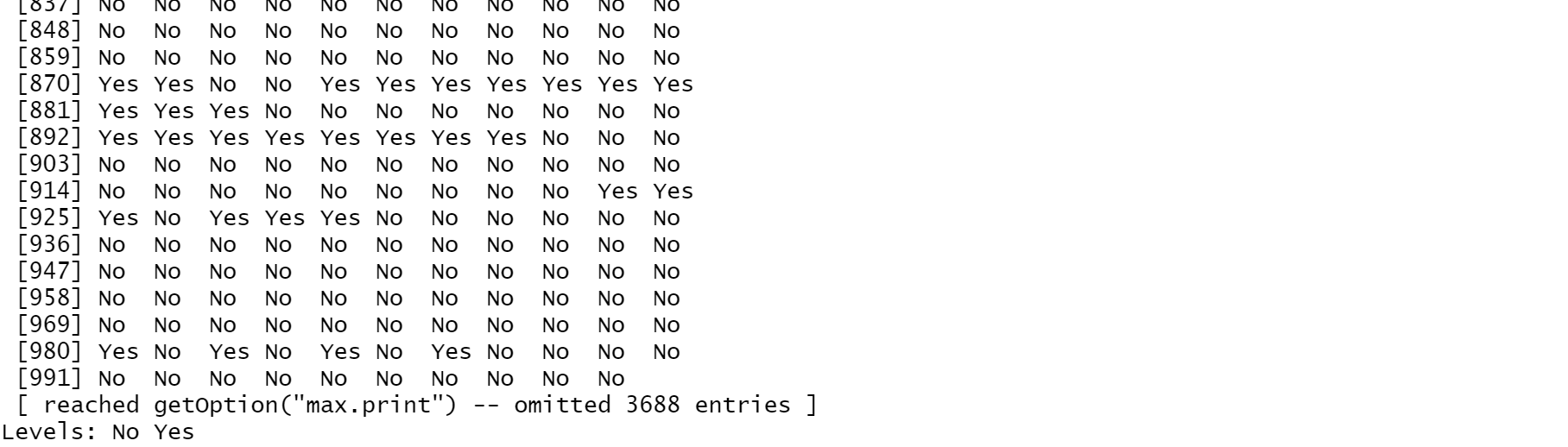








PREDICTING AND CALCULATING ACCURACY OF NAÏVE BAYES MODEL USING CONFUSION MATRIX

**Accuracy of the model is 88.55%.**

The model achieved 88.55% accuracy with a p-value of less than 1. With Sensitivity, Specificity, and Balanced accuracy, the model build is good.

5.4. Findings

|  |  |
| --- | --- |
| **MODEL** | **ACCURACY** |
| FORWARD LOGISTIC MODEL | 91.19% |
| KNN CLASSIFICATION MODEL | 83.85% |
| NAÏVE-BAYES MODEL | 88.55% |

**Hence, Forward Logistic Regression model has the best accuracy out of all the above three models with 91.19%.**