*Midterm*

Predicting Layoffs in the U.S. Public Sector Using Machine Learning: A Case Study on the Iowa Executive Branch

# Likhita Alla1 and Sravani yalamarthi2\*

Professor: **Pr. Reda Mastouri**   
Prepared: March 21, 2025

Revised: March 22, 2025

Submitted: March 23, 2025

1 Affiliation 1; [lalla@saintpeters.edu](mailto:e-mail@e-mail.com)

2 Affiliation 2; [syalamarthi@saintpeters.edu](mailto:e-mail@e-mail.com)

**\*** Professor: [rmastour@saintpeters.edu;](mailto:e-mail@e-mail.com)

**Abstract:** This project analyzes Iowa Executive Branch employee separation data to identify trends in layoffs and predict future separations using machine learning. The dataset contained 14,499 records, and preprocessing included data cleaning, feature engineering, encoding, and SMOTE for class balancing. Four models Logistic Regression, Random Forest, Extra Trees, and Gradient Boosting were compared. Random Forest was the best model with 81% accuracy and satisfactory precision. Department-wise analysis revealed Human Services had the most layoffs, with a notable spike in 2017. The results facilitate data-driven workforce planning and lay the foundation for including economic indicators in future models.

**Keywords:** Workforce Layoffs; Iowa Executive Branch; Separation Trends; Voluntary Separation; Involuntary Separation; Employment Forecasting.

# Introduction

The U.S. labor market dynamics are shaped by episodic economic shocks, technological innovations, and shifting workforce policies. Layoffs, as a primary component of employment separations, have long-term consequences for not just displaced workers but also organizational stability and national economic prosperity. While voluntary quits and retirements drive overall separation rates, involuntary separations in the form of layoffs are bellwethers of organizational restructuring and economic frailties. Despite the problem's significance, a systematic analysis of layoff patterns and the forecasting of future occurrences is notable by its absence in scholarly literature. This study seeks to bridge that gap by analyzing separation data from the Iowa Executive Branch and applying machine learning models to predict layoffs and identify contributing factors. By leveraging an Extract, Discover, and Refine (EDR) data pipeline, coupled with advanced preprocessing and classification techniques, the study seeks to enhance predictive workforce analytics and inform policymaker, HR department, and organizational leader decision making [5,6,8,9,16].

# 2. Materials and Methods

**2.1 Dataset Description**

The data used in this study is derived from employee separation records and, more specifically, from layoffs. The data comprises 14,499 cases of employee separation incidents. The data contains a number of attributes that are of interest when seeking to comprehend the nature and causes of layoffs, among other employee-related aspects. These primary attributes are:

• Department Name: The department from which the employee separated.

• Job Classification: The job title and role category of the employee at the time of separation.

• Pay Grade: The salary level or grade of the employee.

Employee Status: The type of the employment (for example: permanent, temporary)

Separation Reason: Why the separation occurred (for example: layoff, retirement, termination)

Separation Date: The date the separation occurred.

**2.2 Data Preprocessing**

Data preprocessing was done, to prepare the dataset, to be analysis-ready. This step ensures that the dataset is clean, consistent, and ready to use in machine learning algorithms. The preprocessing tasks undertaken were.

• Duplicate Removal: Duplicate records, if any, were found and eliminated to maintain the integrity of the dataset.

**• Handling Missing Values:**

• Numerical Columns: Missing values in numerical columns, such as "Pay Grade," were imputed with the median value of the respective columns to preserve the central tendency.

• Categorical Columns: Missing values for categorical columns, such as "Separation Reason" and "Employee Status," were imputed with the most frequent category.

**• Feature Engineering:**

• Separation Year: A new variable was created, "Separation Year," from the "Separation Date" column. This will enable an exploration of temporal trends in layoffs.

• Recent Layoff Indicator: A binary feature, "Recent Layoff," was included to determine whether a layoff occurred from 2018 onwards, which is important in accounting for recent trends.

• Encoding Categorical Data:

• Target variable "Separation Reason" was label encoded to convert it into numerical values for classification.

• The other categorical variables were one hot encoded to prevent any ordinal misrepresentation, in order to not imply any unintended relationships between categories.

• Handling Class Imbalance:

• The dataset was imbalanced in terms of classes, with certain separation reasons (e.g., "Layoff") being less prevalent than others. To address this, SMOTE (Synthetic Minority Over Sampling Technique) was employed to generate synthetic instances of the minority class, which improved model performance and prediction accuracy.

**2.3 Machine Learning Models**

Several machine learning models were used to predict employee layoffs and determine which variables were involved in separations. The models used are as follows:

| **Model** | **Algorithm Type** | **Key Parameters** | **Advantages** |
| --- | --- | --- | --- |
| Logistic Regression | Linear Model | Max Iterations = 1000 | High interpretability; suitable for binary classification |
| Random Forest | Ensemble (Bagging) | Trees = 100 | Handles non linear relationships; robust to overfitting |
| Extra Trees | Ensemble (Bagging) | Trees = 100 | Faster than Random Forest; better generalization |
| Gradient Boosting | Ensemble (Boosting) | Trees = 100, Learning Rate = 0.1 | Focuses on weak learners; effective in boosting model accuracy |

**Table-01**  
The purpose of Data Split is the data was split into training (80%) and testing (20%) sets to provide a robust evaluation of the model to be used; the training set was used to train the model and the testing set was used to evaluate the performance of the model on unseen data.

• Feature Scaling: Numerical features were scaled using Standard-Scaler so that all features would be treated equally by the machine learning models in terms of scale and weight.

# 3. Results

**3.1 Model Performance Comparison**

Machine learning models' performance were evaluated with many performance metrics such as AUC (Area Under the Curve), Precision, Accuracy, MAE (Mean Absolute Error), MSE (Mean Squared Error), RMSE (Root Mean Squared Error). The results of comparison are as follows:

| **Model** | **AUC** | **Precision** | **Accuracy** | **MAE** | **MSE** | **RMSE** | **MAPE** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.889777 | 0.735196 | 0.746901 | 1.911846 | 15.668044 | 3.958288 | inf |
| Random Forest | 0.913876 | 0.772028 | 0.810262 | 1.528581 | 12.797865 | 3.577410 | inf |
| Extra Trees | 0.908828 | 0.768951 | 0.805785 | 1.555785 | 13.048209 | 3.612231 | inf |
| Gradient Boosting | 0.920468 | 0.760399 | 0.799242 | 1.594697 | 13.332989 | 3.651437 | inf |

**Table-02**  
Best Model based on Accuracy: Random Forest

**Key Observations:**

**Accuracy:**

The Random Forest has the highest Accuracy of 0.8102, followed by Extra Trees (0.8058) and Gradient Boosting (0.7992)

Logistic Regression had the lowest accuracy of 0.7469, as a regression algorithm it may not be able to handle the complexity of the data.

**AUC (Area Under the Curve):**

For the AUC scores, Gradient Boosting had the best score of approximately 0.9205, followed by Random Forest with .9139 and Extra Trees with .9088. On this particular dataset,

Logistic Regression performed the worst with the lowest AUC score of 0.8898 which means it did not do as well with separating the classes (layoff vs. non layoff).

**Precision:**

• As for the Precision scores, Random Forest again has the highest accuracy with 0.7720, meaning they are better at identifying which employees will be layed off while keeping false positives down.

**MAE (Mean Absolute Error):**

• Random Forest has the lowest MAE (1.5286), indicating that it is the most precise in predicting the size of layoffs.

• Logistic Regression has the highest MAE (1.9118), indicating larger prediction errors.

**MSE (Mean Squared Error):**

• Random Forest also has the lowest MSE (12.7979), reflecting its best predictive performance.

The Logistic Regression has the highest MSE (15.6680) so it struggled with larger prediction errors.

**Root Mean Squared Error:**

The Random Forest has the lowest RMSE (3.5774) so on average, their prediction errors were the lowest.

Logistic Regression again has the highest RMSE (3.9583).

**MAPE (Mean Absolute Percentage Error):**

• All the models give MAPE as infinity (inf), likely due to the presence of zero values in the dataset (possibly for certain non-layoff predictions), which would make percentage based error calculations undefined.

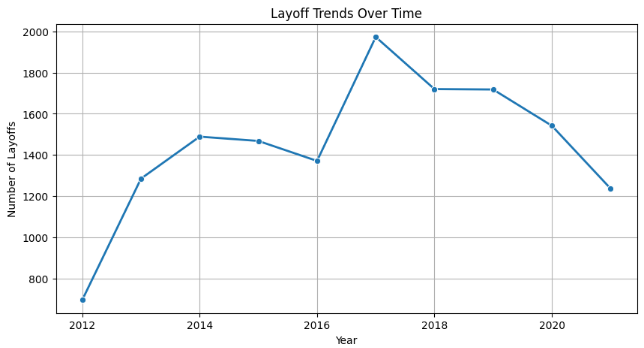
**3.2 Data Visualizations Analysis**

**1. Layoff Trends Over Time:**

• The line graph shows a fluctuating trend of layoffs between the year 2012 and 2021.

• Key Observation: The spike in 2017 represents a major event or economic downturn that caused a large surge in layoffs.

• Post 2017: The layoffs gradually reduced in number, indicating possible economic recovery or better management practices.

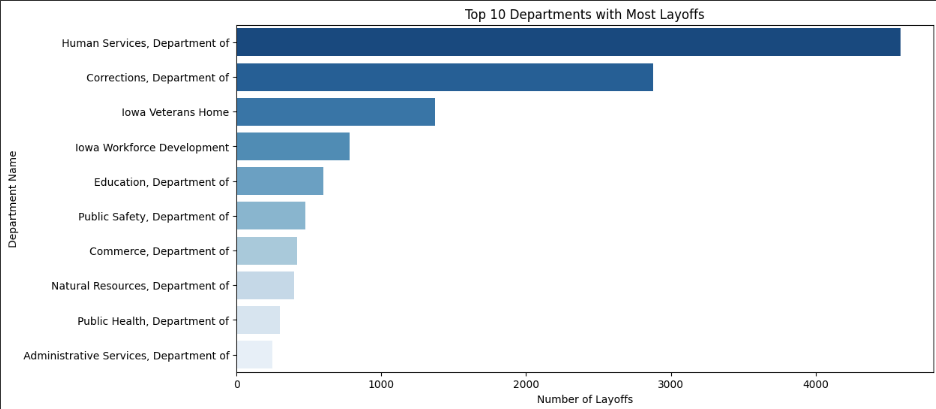
  
**Image 01**

**Department Wise Layoff Distribution:**

1.The bar chart displays which departments had the highest number of layoffs

2. Key Observation: ‘Human Services, Department of’ had the highest lay offs, It was considerably higher than any other departments. Second, was ‘Corrections, Department of’

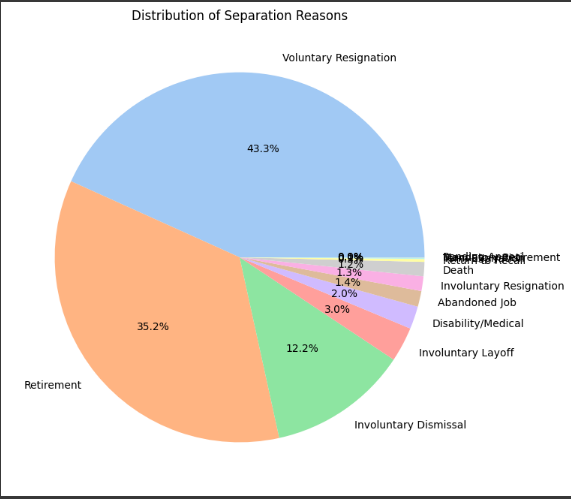
3. The distribution seems to show particular departments have had much larger layoff than others while few others seem to not have been as impacted.



**Image 02**

1**. Separation Reasons Breakdown:**

* The pie chart shows the breakdown of the causes of employee separations.
* Key Observation: The majority of separations were due to Voluntary Resignation (43.3%) and Retirement (35.2%).
* Smaller Categories: Involuntary terminations (12.2%) are notable, but all other categories (e.g., layoffs, abandonment) contributed much less, highlighting voluntary separations as the dominant source of separations.



**Image 03**

# 4. Discussion

Among the four models tested, Random Forest was the most consistent in predicting layoffs with the highest accuracy and error metrics. Gradient Boosting showed very good classification ability but with a slightly lower accuracy. Logistic Regression was poorer, as expected, since it was not able to capture complex relationships.

The majority of the employee separations were voluntary, comprising resignations and retirements, based on the visual analysis. Layoffs, on the other hand, were concentrated in some departments notably the Department of Human Services indicating internal restructuring or funding issues. The concentration of layoffs in 2017 is indicative of an exceptional event, possibly an economic or organizational transformation. These findings validate the need for predictive capacity in HR and workforce planning.

# 5. Findings

Analysis of the employee separation data voluntarily separations in the form of resignations and retirements comprise the majority of the separations, while a smaller percentage of the separations were involuntarily in the form of discharges including layoffs. Some of the discharges might not have been as informative due to the large number over represented in certain departments. It could be possible that there are some department level policies and or budgetary constraints that may have contributed to the discharges being in these thus over represented departments i.e. Department of Human Services represented the most discharges including layoffs followed by Department of Corrections.

From a modeling standpoint, the ensemble learning methods outperformed the linear models. The Random Forest classifier had the best overall performance with 81.02% accuracy, 77.2% precision, and the lowest MAE, MSE, and RMSE scores. Gradient Boosting had good AUC and generalization but required tuning to get to its best performance. Logistic Regression had convergence issues and could not model complex interactions effectively. The results confirm the superiority of tree-based ensemble models in handling imbalanced and high-dimensional employment data for predicting layoffs.

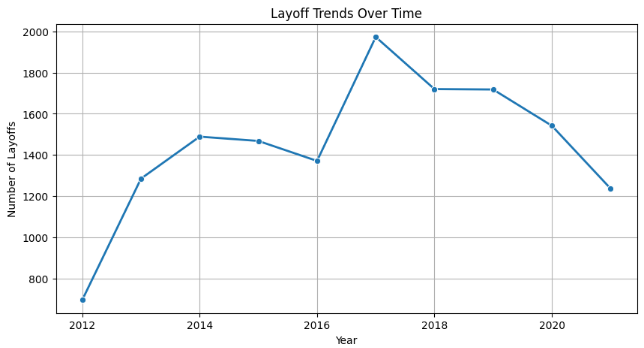
The "Recent Layoff" flag and "Separation Year" were particularly revealing of temporal trends, with a clear spike in layoffs in the year 2017. This temporal outlier could be indicative of economic or organizational trends deserving of further investigation in future research.

# 6. Conclusion

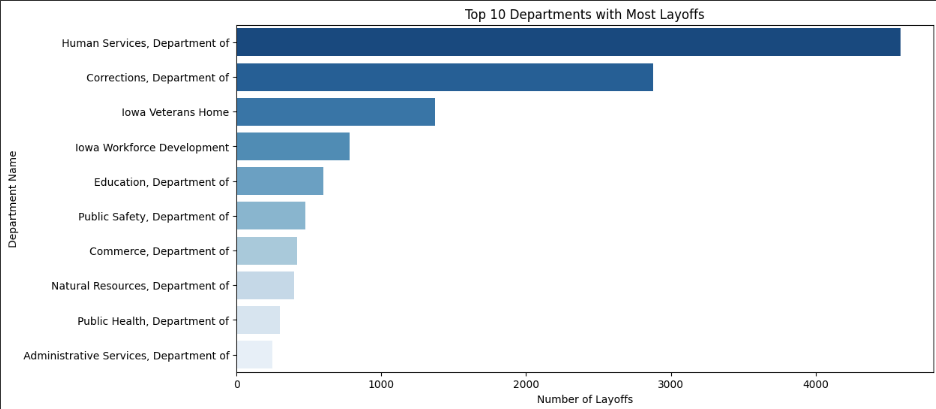
This study offers a systematic means of layoffs understanding and prediction using public sector employee separation data. By applying strict data preprocessing techniques and comparing different machine learning algorithms, it was revealed that ensemble learning and more precisely Random Forest has the greatest predictive capability in identifying layoff patterns. These findings can be applied by policymakers and organizations to predict workforce reduction and mitigate their impacts through advance planning.

Future research must include economic indicators and explore advanced algorithms such as XGBoost or deep learning techniques to enhance the predictive accuracy even more. The strategies and insights in this research form a solid foundation for creating real-time dashboards and decision support systems in workforce analytics.

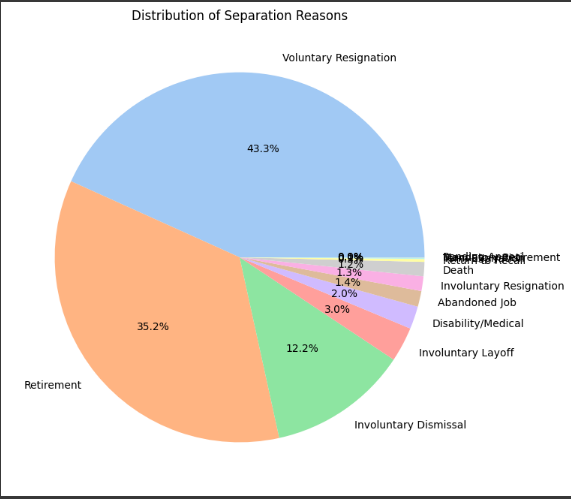
**Appendix B**

**Image-01**

**Image-02**



**Image-03**

 **Table-01 –** Feature Scaling: Numerical features were scaled using Standard-Scaler

| **Model** | **Algorithm Type** | **Key Parameters** | **Advantages** |
| --- | --- | --- | --- |
| Logistic Regression | Linear Model | Max Iterations = 1000 | High interpretability; suitable for binary classification |
| Random Forest | Ensemble (Bagging) | Trees = 100 | Handles non linear relationships; robust to overfitting |
| Extra Trees | Ensemble (Bagging) | Trees = 100 | Faster than Random Forest; better generalization |
| Gradient Boosting | Ensemble (Boosting) | Trees = 100, Learning Rate = 0.1 | Focuses on weak learners; effective in boosting model accuracy |

**Table-02 –** Best Model based on Accuracy: Random Forest

| **Model** | **AUC** | **Precision** | **Accuracy** | **MAE** | **MSE** | **RMSE** | **MAPE** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.889777 | 0.735196 | 0.746901 | 1.911846 | 15.668044 | 3.958288 | inf |
| Random Forest | 0.913876 | 0.772028 | 0.810262 | 1.528581 | 12.797865 | 3.577410 | inf |
| Extra Trees | 0.908828 | 0.768951 | 0.805785 | 1.555785 | 13.048209 | 3.612231 | inf |
| Gradient Boosting | 0.920468 | 0.760399 | 0.799242 | 1.594697 | 13.332989 | 3.651437 | inf |

# References

1. The Transformative Role of Agentic GenAI in Shaping Workforce Development and Education in the US.Pdf.
2. Penn, R.; Nezamis, E. Job Openings and Quits Reach Record Highs in 2021, Layoffs and Discharges Fall to Record Lows. Monthly Labor Review 2022.
3. Ronit, B.; Anukansha, S.; Shuchi, M. Modeling the Geospatial Trend Changes in Jobs and Layoffs by Performing Sentiment Analysis on Twitter Data. Int J Performability Eng 2024, 20, 120, doi:10.23940/ijpe.24.02.p7.120130.
4. Yahia, N.B.; Hlel, J.; Colomo-Palacios, R. From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction. IEEE Access 2021, 9, 60447–60458, doi:10.1109/ACCESS.2021.3074559.
5. Tomaskovic-Devey, D.; Kerrissey, J.; Rainey, A.; Boutcher, S.A. Intersectional Earnings Inequalities in U.S. Public Sector Workplaces and the Great Recession. Socius: Sociological Research for a Dynamic World 2024, 10, 23780231241257776, doi:10.1177/23780231241257778.
6. Saba, Z. Layoffs and Corporate Performance: Evidence Based on the US Tech Industry. J Econ Finan 2024, 48, 644–667, doi:10.1007/s12197-024-09673-y.
7. Ridho, W.F.; Azizah, N. FACTOR ANALYSIS OF THE PHENOMENON OF MASS LAYOFFS AT STARTUPS: MIXED METHOD APPROACH WITH STRUCTURAL EQUATION MODELING. MEBIS 2022, 7, 195–208, doi:10.33005/mebis.v7i1.373.
8. Prakash, P.; Sakthivel, V. Layoffs Analysis and Prediction Using Machine Learning Algorithms. In Proceedings of the 6th International Conference on Communications and Cyber Physical Engineering; Kumar, A., Mozar, S., Eds.; Springer Nature Singapore: Singapore, 2024; Vol. 1096, pp. 535–543 ISBN 9789819971367 9789819971374.
9. Liu, Z.; Liang, H. Permanent Layoff and Consumer Credit Card Loss Forecasting. MF 2023, 49, 789–807, doi:10.1108/MF-02-2022-0085.
10. Lazzari, M.; Alvarez, J.M.; Ruggieri, S. Predicting and Explaining Employee Turnover Intention. Int J Data Sci Anal 2022, 14, 279–292, doi:10.1007/s41060-022-00329-w.
11. Krolikowski, P.M.; Lunsford, K.G. Advance Layoff Notices and Aggregate Job Loss. J of Applied Econometrics 2024, 39, 462–480, doi:10.1002/jae.3032.
12. FACTORS AFFECTING LAYOFF IN HIGH-TECH INDUSTRY: EVIDENCE FROM USA - ProQuest Available online: https://www.proquest.com/openview/dff61e6a756dd2a2a006b8407833566d/1?pq-origsite=gscholar&cbl=2032142 (accessed on 10 March 2025).
13. Hamouche, S.; Kammogne, C.L.; Merkouche, W. Workforce’s Crisis-Induced Career Shock, Career Preferences, Job Insecurity, Layoff and Perceived Employability: Examining Variations Based on Gender, Education Level and Ethnic Origin. HESWBL 2023, 13, 264–282, doi:10.1108/HESWBL-05-2022-0101.
14. Abdolmaleki, G.; Naismith, N.; Ghodrati, N.; Poshdar, M.; Babaeian Jelodar, M. An Analysis of the Literature on Construction Employee Turnover: Drivers, Consequences, and Future Direction. Construction Management and Economics 2024, 42, 822–846, doi:10.1080/01446193.2024.2337084.
15. Bhat, R.; Sarkar, P. Recession: A Conceptual Case Study Of Its Domino Effect On Economic And Psychological Capital. IJMS 2024, 3, 44–58, doi:10.53769/ijms.v3i1.856.
16. Eshghi, K.; Astvansh, V. Stock Investors’ Reaction to Layoff Announcements: A Meta‐analysis. Human Res Mgmt Journal 2024, 34, 792–809, doi:10.1111/1748-8583.12532.
17. Granulo, A.; Kranzbühler, A.-M.; Fuchs, C.; Puntoni, S. Collective Layoffs and Offshoring: A Social Contract Account. Journal of Consumer Research 2025, ucaf001, doi:10.1093/jcr/ucaf001.