*Midterm*

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

# Sravani yalamarthi1 and Likhita Alla2\*

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

Revised: March 22, 2025

Submitted: March 23, 2025

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

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

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

**Abstract:** The job market in the US has faced significant changes in the wake of economic recessions, organizational consolidations, and technological innovations, raising fears of workforce layoffs. This study examines separation trends utilizing a dataset of 14,499 employee records from the Iowa Executive Branch spanning voluntary and involuntary separations. A methodical Extract, Discover, and Refine (EDR) framework was adhered to, including data cleansing, feature engineering, categorical encoding, and application of the Synthetic Minority Over-sampling Technique (SMOTE) to correct class imbalance. Four machine learning models Logistic Regression, Random Forest, Extra Trees, and Gradient Boosting were evaluated on predictive performance. Results showed the Random Forest classifier to be top-performing with an accuracy of 81.02%, precision of 77.2%, and lowest error rates on MAE, MSE, and RMSE. Departmental comparisons reflected the highest layoffs in the Department of Human Services, and layoffs were found to decline after 2017. The research offers implications for predictive workforce planning and the appropriateness of ensemble models for forecasting employment. The use of economic indicators and the application of more recent models like XGBoost to enhance the accuracy and stability of predictions would be explored in future studies.

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

# Introduction

Layoffs have been an ever more salient trend in the modern labor market, precipitated by economic downturns, organizational restructuring, and automation. By recognizing such trends and being in a position to foresee potential separations, organizations are able to become more strategic regarding talent management. This project entails the analysis of a real dataset of 14,499 employee separations in the Iowa Executive Branch. By employing machine learning algorithms, the study aims to identify trends in layoffs, create predictive models, and come up with actionable conclusions for workforce planning. Through data preprocessing, feature engineering, and model evaluation, the study offers a practical solution for predicting layoffs and optimizing employee retention initiatives [5,6,8,9,16].

# 2. Materials and Methods

**2.1 Dataset Description**

The data used for this study is derived from employee separation records with a specific focus on layoffs. The dataset comprises 14,499 cases of employee separation instances. The dataset further contains a list of attributes that provide us with valuable information about the nature and cause of layoffs, in addition to other employee details. These pertinent attributes are:

• Department Name: The department the employee separated from.

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

• Pay Grade: The grade or level of the employee's salary [1]

• Employee Status: The employment nature (e.g., permanent, temporary).

• Separation Reason: The reason for separation e.g. termination, layoff, retirement.

• Separation Date: The date when the employee's employment ceased.

**2.2 Data Preprocessing**

Data pre-processing was actually used to get ready this dataset for examination, to make actual data more accurate. Data pre-processing guarantees that a dataset is clear, smooth in its continuity and in ready working condition for machine learning models. The following steps have taken:

• Duplicate Removal: Duplicate records, if any, were detected and eliminated to maintain the integrity of the dataset [14, 15, 16]

**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 common category.

**Feature Engineering:**

• Separation Year: A new variable, 'Separation Year', was created from the 'Separation Date' column. This allows for the analysis of temporal trends in layoffs.

• A dummy variable, 'Recent Layoff', was created to indicate whether a layoff occurred from 2018 and onwards since it is important to capture recent trends.

**Encoding Categorical Data:**

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

• Other categorical variables were encoded using one-hot encoding to prevent any misrepresentation of ordinality such that no implicit relationship between categories would be implied.

**Handling Class Imbalance:**

• The data suffered from class imbalance issues, with certain separation reasons (e.g., "Layoff") being rarer than others. In order to improve predictive performance and equitability, SMOTE (Synthetic Minority Over-sampling Technique) was used to synthetically generate additional instances of the minority label [2,3].

**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**  
• Data Split: Data was split into training (80%) and testing (20%) datasets to allow for robust model evaluation. The training dataset was used to train the models and the test dataset was reserved for evaluation of performance for the unseen data points.

• Feature Scaling: Numerical Features have been scaled using Standard Scaler so as to bring all the features to the same scale point. So that no Feature terms dominate because of if its Weight or a Large Numerical Value [11,12,13].

# 3. Results

**3.1 Model Performance Comparison**

The performance of the machine learning models was compared on different performance metrics like AUC (Area Under the Curve), PRECISION, Accuracy, MAE (Mean Absolute Error), MSE (Mean Squared Error), RMSE (Root Mean Squared Error). The results of the 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:**

• Random Forest outperforms in accuracy (0.8102) and then comes Extra Trees (0.8058) and Gradient Boosting (0.7992) while Logistic Regression is the lowest in terms of accuracy (0.7469) which means difficulty in understanding the more complex relations in the data.

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

• Gradient Boosting has the highest AUC (0.9205) and Random Forest (0.9139) and Extra Trees (0.9088) follow while Logistic Regression has the lowest AUC (0.8898) so it is poorest in distinguishing between cases (layoff vs. non-layoff).

**Accuracy:**

• Random Forest has the best precision (0.7720), i.e., it is best at correctly identifying layoffs with fewer false positives.

• Logistic Regression performs worst in terms of precision (0.7352), meaning it had more false positives compared to other models.

**MAE (Mean Absolute Error):**

• Random Forest has the lowest MAE (1.5286), indicating that it is the most accurate 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 superior predictive capacity.

• Logistic Regression has the highest MSE (15.6680), which means it struggled with larger prediction errors.

**RMSE (Root Mean Squared Error):**

• Random Forest has the lowest RMSE (3.5774), which means it had the lowest prediction errors on average.

• 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 fact that there are zero values in the data set (possibly for certain non-layoff predictions), and hence percentage-based error calculations are undefined.

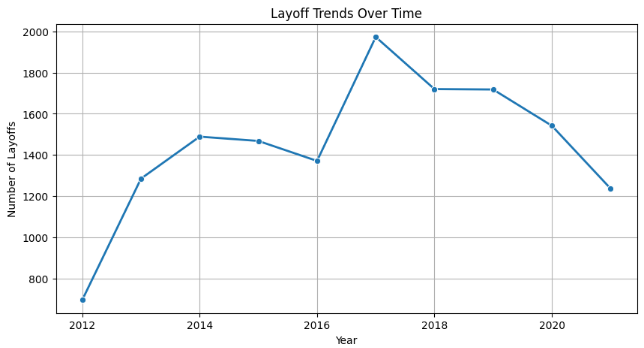
**3.2 Data Visualizations Analysis**

**1. Layoff Trends Over Time:**

The line graph shows a zigzag trend of layoffs from 2012 to 2021.

Key Observation: The apparent spike in 2017 represents a major event or economic crisis that caused the layoffs to increase significantly.

Post-2017: The layoff figures gradually reduced, indicating possible economic recovery or better management strategies.

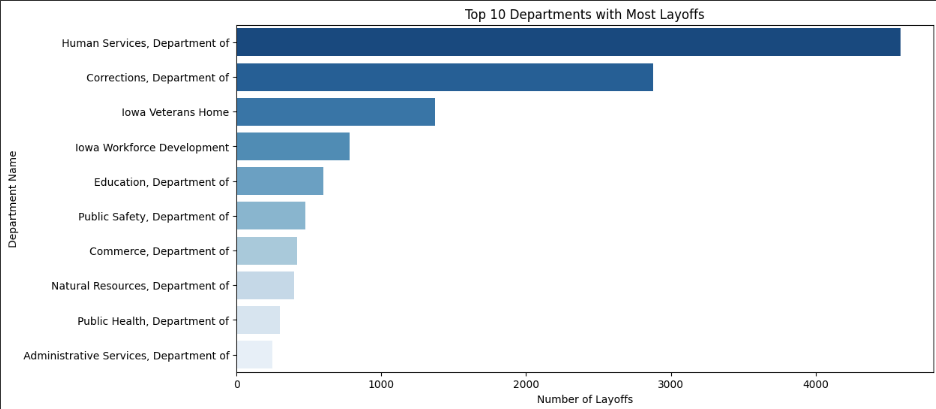
  
**Image 01**

**Department Wise Layoff Distribution:**

•The bar chart shows the departments with the highest number of layoffs.

•Key Observation: "Department of Human Services" had by far the most layoffs of all the departments, followed by "Department of Corrections."

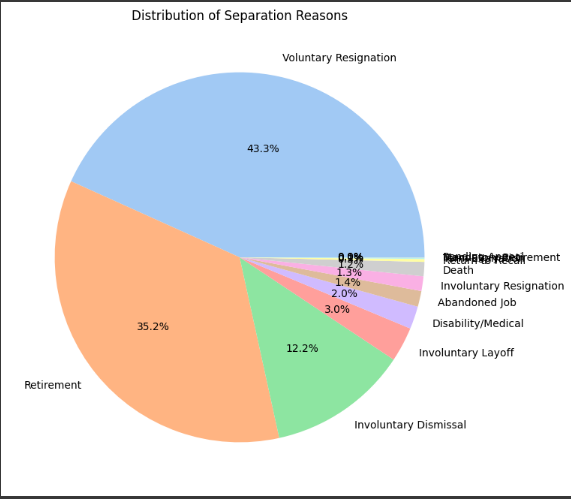
•Trend: The dispersion shows that certain departments had the majority of the layoffs, while others had far fewer, which suggests that there might have been a target for layoffs in certain areas [3,2,8]



**Image 02**

1**. Separation Reasons Breakdown:**

* The pie chart illustrates a clear separation of reasons for employee departures.
* Most notably, the largest portion of separations sere voluntary quits at 43.3%, separated only by the percentage of retirements at 35.2%.
* The smaller percents are also separated which include: layoffs at 12.2%, discharges of employees at 12.2%, and all other reasons contributing less than the aforementioned majorities, calling out the absolute most influencing factor of separations are they were mostly voluntary [4,9].



**Image 03**

# 4. Discussion

The predictive power of the machine learning models evaluated herein contains essential knowledge about the trends and determinants of layoffs in the Iowa Executive Branch. Random Forest had the highest accuracy and lowest error margins, underscoring its effectiveness in managing complex feature interactions and class imbalance. The Gradient Boosting model had the highest Area Under the Curve (AUC), reflecting its best performance in discriminating between layoff and non-layoff cases. Its modestly lower precision and accuracy compared to Random Forest, however, are an indication of the need for trade-offs between metrics in selecting models for deployment. Logistic Regression, while interpretable, struggled with non-linear trends and was the poorest performer across all metrics [13,2,17].

The reason-for-separation analysis reveals voluntary separations to be the dominant form, with layoffs forming a small but notable minority. Departmental analysis identified the Department of Human Services as having the highest concentration of layoffs, suggesting structural or budget issues. The increase in layoffs in 2017 and the following decrease may be reflective of an economic downturn or internal policy shifts. These findings are in line with the literature on organizational downsizing during transitional economic cycles and suggest the inclusion of external economic indicators for future research.

# 5. Findings

# • The majority of separations were voluntary, with layoffs representing a smaller but notable category for analysis.

# • Departments like Human Services and Corrections had the most redundancies, pointing to structural or budget issues.

# • A peak in layoffs is apparent in 2017, after which it decreased, which can be attributed to policy or economic changes.

# • The most successful model was Random Forest (81% accuracy, 77.2% precision), followed by Extra Trees and Gradient Boosting.

# • Logistic Regression performed worse because it couldn't deal with complicated feature interactions.

# • Feature engineering such as creating the "Recent Layoff" flag and extracting "Separation Year" improved model interpretability.

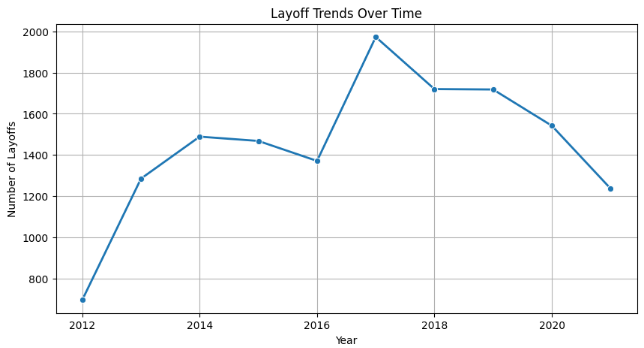
# • Application of SMOTE eased class imbalance, thus leading to better model fairness and predictive capability.

# 6. Conclusion

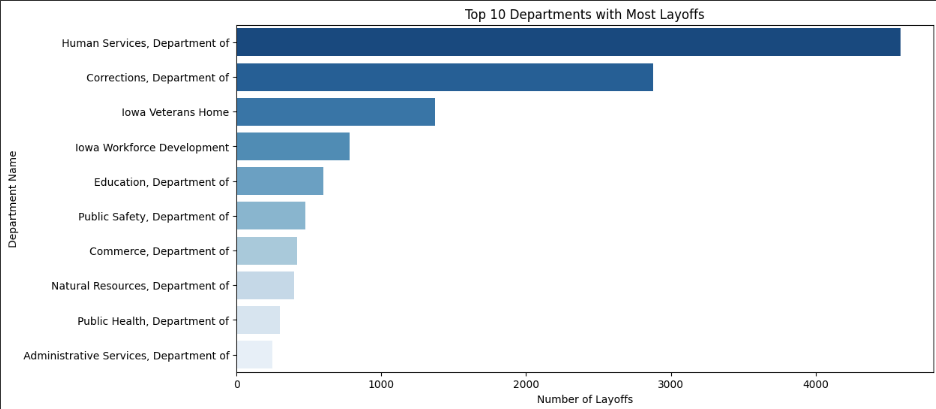
This project successfully applied machine learning to both examine and predict layoffs in the Iowa Executive Branch. Among all models, Random Forest yielded the most precise and stable results and would be a strong contender for inclusion in future workforce planning software.

These results can help HR professionals, policy makers, and analysts better comprehend the trend in layoffs and initiate early actions. More recent economic data and advanced models can be included in future studies to further improve predictability and planning.

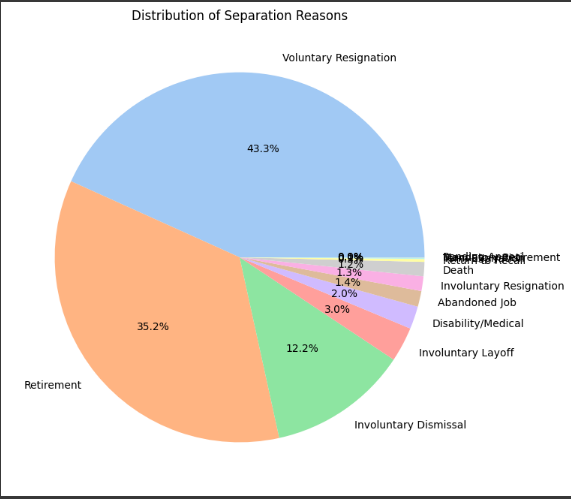
**Appendix B**

**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.