**[Employee Attrition & Performance Prediction Model](https://drive.google.com/drive/folders/1G-W2jLvK2AcmyCS2IQP4fUvaa-ypugDG?usp=sharing)**

[**Python Code**](https://drive.google.com/file/d/12hkxsgHMQh79wsIiRNlBMeQyXO_M8HBi/view?usp=sharing)

[**Github\_Repository**](https://github.com/xaltyPasta/Employee_Attrition_Prediction)

**1. Introduction**

This Assignment Solution presents Logistic Regression and Random Forest machine learning algorithms for predicting employee attrition. The dataset used is from a multinational corporation. The Logistic Regression model provides interpretable results and identifies key factors contributing to attrition. The Random Forest model, while less interpretable, offers superior predictive accuracy. Both models were evaluated using cross-validation. Additionally, hyperparameter tuning was applied using GridSearchCv, including the Random Forest model, to further enhance their performance.

**2. Methodology**

**2.1 Data Preprocessing**

**2.1.1. Dataset Cleaning**

We addressed missing values, removed duplicates, and corrected data types in our dataset  
using pandas library :  
**df.isnull().sum()  
df.dropna()**

**2.1.2. Feature Analysis and Removing Duplicates:**

Feature analysis involves identifying the most influential variables Removing duplicates ensures the dataset's integrity, eliminating redundant and irrelevant entries for accurate prediction.

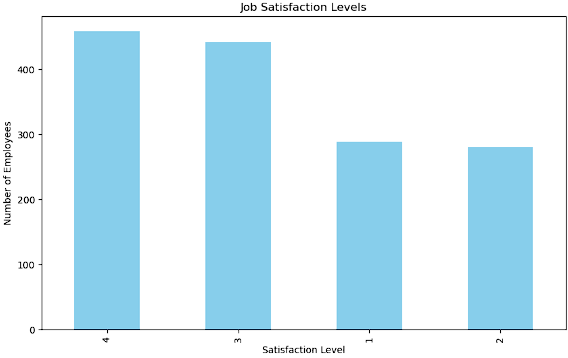
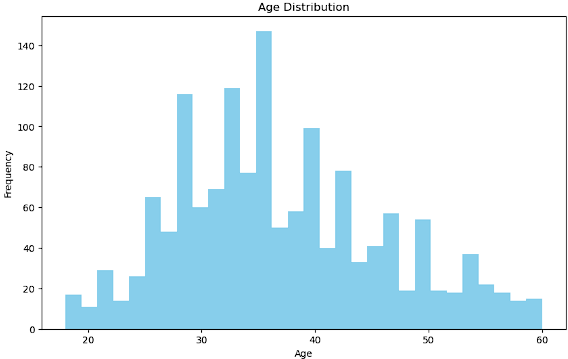
**2.2 Encoding of Objects**

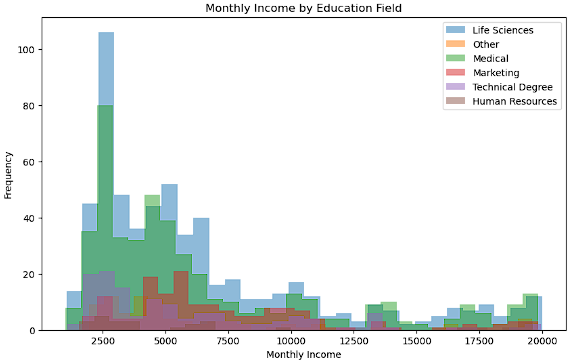
We performed encoding of object type columns using pandas and **Labelencoder** :  
**pd.getdummies()  
Labelencoder()**  
And dropped the object columns after encoding.

**3. DataSet Visualization:**

we plotted different graphs using matplotlib based on to help in analysing and feature selection :

* **age distribution**
* **Job Satisfaction**
* **Monthly Income by Education Field**





**4. Evaluation & Results:**

We experimented with two classification algorithms :

* **Logistic Regression**
* **Random Forest**

**4.1)Logistic Regression:**

Relying on the simplicity and interpretability of Logistic Regression, we utilized this classical statistical model to predict binary outcomes based on a set of independent variables. Despite its simplicity, Logistic Regression proved to be a formidable tool, particularly effective when the classes exhibit linear separability, providing valuable insights.

**Evaluation:**

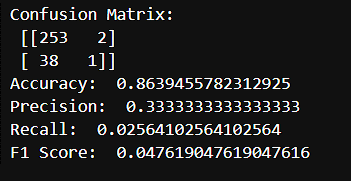
* **Mean\_Squared\_Error: 0.1360544217687075**
* **Cross\_Validation\_Score:**

**[0.80932203 0.82553191 0.82553191 0.82978723 0.83404255]**

**Average cross-validation score: 0.8248431301839163**



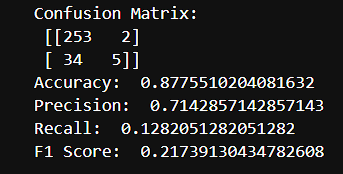
**Metrics Like Accuracy, Precision, Recall ,and F1 Score with the help of Confusion matrix:**



**4.2)Random Forest:**

Random Forest is a versatile machine learning algorithm capable of performing both regression and classification tasks. It’s known for its simplicity and ability to handle large datasets with high dimensionality. Here are some key features and advantages like: Ensemble method, handling missing values, outlier detection high accuracy etc.

**Evaluation:**

**Metrics Like Accuracy, Precision, Recall ,and F1 Score with the help of Confusion matrix:**

**HyperParameter Tuning**: ***using GridSearchCv:***

**Best parameters found:**

Best Parameters: {'max\_depth': 6, 'max\_features': 'sqrt', 'min\_samples\_leaf:2,'min\_samples\_split': 5, 'n\_estimators': 100}

***So that we can improve and Optimize model’s performance by using these parameters .***

**6. Findings an insights:**

**Age:** There might be a certain age group that’s more likely to leave the company. For example, younger employees might be more prone to attrition due to a desire for varied experiences and growth opportunities.

**BusinessTravel:** Employees who travel frequently might have higher attrition due to the stress and lifestyle challenges associated with frequent travel.

Department: Some departments might have higher attrition rates due to the nature of the work or the department culture.

**EducationField & Education:** The field of education and the level of education might influence attrition. For instance, employees from certain fields might be more sought after and hence more likely to leave.

**JobSatisfaction & EnvironmentSatisfaction:** Low job satisfaction and poor work environment could be strong predictors of attrition.

**MonthlyIncome:** Employees with lower income levels might be more likely to leave the organization in search of better pay elsewhere.

***Recommendations:***

**Improve Work Conditions:** Improving the work environment and job satisfaction could help reduce attrition. This could involve measures like improving workplace facilities or addressing employee grievances effectively.

**Competitive Pay:** Ensuring that the pay is competitive as per the industry standards could help retain employees.

**Career Growth Opportunities:** Providing clear career growth paths and opportunities for upskilling can help in retaining employees, especially the younger ones.

**Work-Life Balance:** For employees who travel frequently, ensuring they get adequate time off and support for a healthy work-life balance can help reduce attrition.

**7. Conclusion:***Given the absence of time-series data in this dataset, all classification tasks have been designed to train a machine learning model on a subset of the employees and test it on the remaining unseen employees. This approach allows us to evaluate the generalizability of our conclusions and assumptions to unseen data. If these models perform well on the unseen data, we can confidently extrapolate our conclusions to the entire employee population, knowing that our predictions will generally hold true. This methodology underscores the importance of robust model validation techniques in predictive analytics.*