**HR Analytics project- Understanding attrition in HR**

1. **Problem Definition**

Every year companies invest significant resources in hiring and training employees ,aiming to enhance their workforce’s effectiveness and productivity. However, attrition remains

persistent challenge for HR departments across industries. Employee turnover not only

disrupts operations but also incurs substantial costs associated with recruitment ,training

and loss of productivity . HR analytics offers a strategic approach to address this issue by leveraging

data –driven insights to understand and mitigate attrition.

The primary objective of this HR analytics is to identify factors contributing to employee attrition

And develop e predictive models to anticipate and mitigate it effectively . While enhancing employee performances is one aspect of HR analytics .The focus here is specifically on understanding and reducing attrition rates within an organization.

**2 . Data Analysis**

The dataset for this project comprises various employee –related attributes such as Age ,

Department,distance from home, education, environment satisfaction, gender, job involvement,

Job level, job role, job satisfaction, marital status, monthly income, number of companies worked,

Overtime, percent salary hike, performance rating, relationship satisfaction, total working years,

Training time last year, work life balance, years at company, years in current role , years in last promotion, years with current manager etc..

Dataset contains 1470 rows and 31 columns.All column count are equal,no missing values .

Dataset contains categorical and continuous data.Target variable is categorical so its classification

Problem model building . Target column value counts ,we can see some class imbalance

Exploratory data analysis (EDA ) techniques are employed to gain insights into the underlying

Patterns and trends with in the data

**3.EDA concluding remarks**

Based on EDA ,several insights emerge regarding the relationship between various factors and

Attrition(target).For instance ‘Monthly income’ increases Attrition count is less. Attrition on department wise we can see that Research & Development department have less attrition count

Comparing marital status with Attrition ,Divorced attrition count is less .

**4. pre processing pipeline**

Checking distribution of continuous data ,we can find some outliers present .we confirm it with boxplot. removed outliers using zscore .we can find skewness for ‘monthly income’. reduced skewness using cube root method. Dataset contains some categorical column we have to convert

Categorical data to continuous data using some encoding techniques. Here we use Ordinal encoder to convert categorical data to continuous data. Visualizing correlation between features and target

Using Heatmap . There is no features have high correlation with target variable, but target variables slightly influenced by some features .Divide dataset into features and target . scaled the features ata using Standard Scaler .checkd multicolinearity of features using variance inflation factor. We can see that no features have multi co linearity issue . Here we can see the data is not balanced hence its classification problem we will balance the data using oversampling method

**5. Building machine learning models**

Several Machine learning algorithms are employed to develop predictive models for employee

Attrition. These include Logistic Regression, Gradient Boosting Classifier,RandomForestRegression

SVC,AdaBoostClassifier,BaggingClassifier,ExtraTreesClassifier,DecisionTreeClassifier and

KNeighborsClassifier. Each algorithm evaluated based on performance metrics such as accuracy score ,Precision, recall,and F1 score. We have done cross validation to check the model is generalized ,not overfiting and underfiting .after cross validation we can see that KNN model is the best performing model among the models we build because the deference between accuracy score and cross validation score is less in KNN model..Then we have done hyperparameter tuning for KNN model to

Get better score .using GridSearchCV to find best parameters .Before hyper parameter tuning accuracy score was 77, after hyper parameter tuning accuracy score became 88 .

**6.Concluding remarks**

In Conclusion HR Analytics offers valuable insights that can drive strategic decision making with in

Organization. By leveraging data driven approaches ,HR departments can identify potential areas of

Improvements , devise targeted interventions to enhance employee satisfaction and retention and

Ultimately foster a more productive and engaged workforce .However,it is essential to recognize the

Ethical implications of utilizing employee data and ensure compliance with privacy regulations and ethical standards.

Through this project we have demonstrated efficacy of HR analytics in predicting employee attrition

And provided a roadmap for implementing similar initiatives in other organizations .As the field continues to evolve , embracing data analytics in HR practices will undoubtedly become increasingly vital for organizational success in the digital age .