**Blog/Article on HR Analytics Project**

**Problem Definition**

In today's competitive business landscape, organizations are increasingly turning to data-driven approaches to gain insights into their workforce and make informed decisions. Human Resources (HR) analytics, also known as people analytics, is a powerful tool that leverages data science techniques to analyze HR data and derive actionable insights. One common application of HR analytics is in predicting and managing employee attrition, which can have significant implications for organizational performance and success. In this article, we will explore the concept of HR analytics, delve into the intricacies of attrition prediction projects, and discuss best practices for implementing such projects effectively.

As in the 4th Phase of the Evaluation Project, I select Hr analytics project to Explain the learnings of the same project, I got the dataset from the Flip Robo Technologies and with out any web scrapping task I move forward for the rest of the steps to explore my project there were a lot of feature (Dependent and independent) present in this dataset my target was to check attrition of the employes working in the company.

**Understanding HR Analytics**

HR analytics involves the application of statistical analysis, data mining, and machine learning techniques to HR data to identify patterns, trends, and relationships. By analyzing various HR metrics such as employee demographics, performance ratings, compensation, engagement surveys, and exit interviews, organizations can gain valuable insights into their workforce dynamics and make data-driven decisions to improve employee retention, productivity, and satisfaction.

First of all in my Jupyter note book I import all the important libraires such as pandas, numpy,matplot and seaborn for data processing . Then in the next step I imported the csv file of the dataset into the python notebook using pandas read\_cvs method , after importing the file i displayed top and the bottom rows of the dataset to analyse and check the features of the dataset, also in this phase I can see all the features present in the dataset whether dependent or independent, also check the type of values present in the dataset, then I moved on next step for further study.

Attrition prediction is a key focus area within HR analytics, as employee turnover can have significant financial and organizational implications. Attrition prediction projects aim to develop models that can forecast the likelihood of employees leaving the organization within a certain time frame. These models typically utilize historical HR data, including employee demographics, job characteristics, performance metrics, and tenure, to predict attrition risk factors and identify at-risk employees.

Then I displayed all the rows and columns present in the dataset using shape() method of the pandas dataframe, after this step I also displayed the full dataframe information such as ferature names, null values, and data types of all the features using pandas info() method, with this step I got full information about the dataframe mostly datatypes of the dataset mostly datatypes are int, float and object, int and float represents the continuous values while as the object represent the categorical values.

**Data Analysis**

After getting all the information of the dataframe I checked the null values present in the dataset using isnull() method , but in this project there where was not any null values present , in case of null values I would go for null values treatment like filling null values using fillna() method of pandas library with the mean, median, mode values, but being careful while filling the null values it is important to take care of outliers preset in the features, if outliers are present then fill null values using appropriate method.

Then in the next step I display number of values present in the particular columns using the pandas value\_counts method, till now I observer all the feature details of the data set.

Feature engineering is the science of features of the data that we are analyzing and getting the appropriate knowledge of the features, also do the required operations in order to make the assumption in a smooth manner. And also checking the correlation between the features in which we can see that how the features are positively and negatively related to each other.

In this feature engineering also we are getting the statistical information of the data frame to check the minimum and maximum value in each of the numerical column and also checking the quartiles of the dataset. After completing all these steps the I move forward to the Exploratory Data Analysis(EDA).

**Exploratory Dat Analysis**

In the HR Analytics project I started to visualize the feature values graphically, in the step of data analysis mostly matplot, and seaborn library are used to crate the graphs and charts. Firstly I plotted graph of score of Attrition employees from the company, there are only two values (yes, no ) and in this project there are less number of Attrition.

Now after this step my step was to analyze attrition rate with every single feature of which attrition is dependent, checking the which department is mostly churning in the project Research and Development has the high churning rate, after this checking attrition vs gender and males has the high churning rate, checking the marital Status vs Attrition and married one has the high Attrition rate, by these analysis I can say the mostly married males are mostly churning the company, also we cane make the assumption on the rest of the fields on the data set for better results.

Then I started to create scatter plot to check the effect and relation of different fields in order analyze the full summary of the dataset. Also boxplot is used to check the outliers present in the fields of the dataset,

Checking the distribution of the features present In the dataset Is very useful to analyze the data. The distribution plot show the curve of the data there are a number of distributions such as Normal Distribution, log Normal distribution, and power law distribution. By checking the distribution we can convert the data into the normal distribution for the better results.

After this step I go for the data encoding of the categorical columns using sklearn, and I used Ordinal Encoding for the same process after converting the categorical into numerical values then move to to the next process.

**Creating Train Test Split**

After all the steps then I separate the dependent and independent feature also I dropped the feature which are not necessary for the analysis.

Before training the models always remember that we need to create training and testing data to train the models , machine learning model are good to process the numerical data , also the machine learning model are more accurate on the numerical data. By the use of train test split training set and the testing set is created namely x\_tran, x\_test,y\_train,y\_test, in which x\_train contain the number of independent features while as x\_test contains the test part of the independent variables, also y\_train contains the training data of the dependent or target variable and y\_test contains the test data of the target variable , it is important to set the test size while using train test split.

**Building The Models**

The next step is to train and import the necessary model using training data, also we need to take care of one thing if the target variable has the continuous values or the categorical values, if the target variable has categorical then we need to use the classification models, if the target variable contains the continuous values the we need to use regression modes.

there are various models which I use to predict the attrition rate of the employees in this project, such as LogisticRegression, SVC, RandomFrorest Classifier and Gradient Boosting Classifier. After training all the models using training data, then predict the result using the testing data.

After all this, it is important to check the accuracy score of the models using sklearn.metrics, there are mainly four scores used in this HR Analytics project accuracy\_score,precision\_score,recall\_score,f1\_score. After checking the score of all the models then I started to visualize the feature importance of the models.

Finally, I saved the best model using joblib dump method, and in made a new dataframe with the new values to predict the result of the saved model.

Conclusion:

HR analytics, particularly attrition projects, holds immense potential for organizations to proactively manage their workforce and mitigate talent loss. By leveraging data-driven insights organizations can identify risk factors, implement targeted retention strategies, and foster a culture of employee engagement and satisfaction. As organizations continue to embrace HR analytics as a strategic imperative, the ability to predict and manage attrition will become increasingly critical in driving organizational success and comparative advantage in the modern business landscape.