**Document on elaboration of my project and how I used to test my project**

**I** would like to showcase a project that I have completed Called**: Employee Turnover Prediction**

**a. What did the System Used to do ?**

The project employee Turnover is a machine learning project by using python .This project has been done to predict the employ turnover.

Employee turnover refers to the percentage of workers who leave an organization and are replaced by new employees .It is one of the most significant problems an organization can encounter throughout its lifecycle, as it is difficult to predict and often introduces noticeable voids in an organization’s skilled workforce

One way organizations deal with this problem is by predicting the risk of attrition of employees using machine learning techniques thus giving organizations leaders and Human Resources (HR) the foresight to take pro-active action for retention or plan for succession.

We will work on simulated HR data from kaggle to build a classifier that helps us predict what kind of employees will be more likely to leave given some attributes

The main objective of this project is to predict the percentage of employees who leave from a company

As a result, it is imperative that organizations formulate proper recruitment, acquisition and retention strategies and implement effective mechanisms to prevent and diminish employee turnover, while understanding its underlying, root causes.

**b. What other systems have you seen in the wild like that?**

The Employee turnover was often predicted by the HR department based on the experience. This practice leads to inaccurate results that is recruitment of required number of employees is uneven

Calculations and predictions made through the use of predictive analytics are based on factual and historical data; that data is used to assess what's likely to happen in the future based on what's happened in the past

Employee turnover rates play heavily into the way most organizations go about day-to-day business. Everything from the hiring process to the floor plan of an office to the way that management commits to treating staff can be influenced by how likely you feel you are to lose employees

Among many other reasons, this is because a whole field emerged from the idea of using data to support human resources: HR analytics (also called people analytics) is about changing the way of recruiting and retaining talent based on data-driven insights (Isson& Harriot, 2016). This way data analytics are used to predict behavioural patterns (e.g., attrition rates, training costs, productivity) which are inherently informative to the respective management because it can guide their decision-making process. Based on the successful implementation of machine learning algorithms, some of the big players already apply predictive analytics to decrease attrition and increase retention of their profitable employees.

**c. How do you approach the testing problem?**

For the testing problem i have built few test cases to perform the testing. It is quite hard to built test cases for this type projects because it runs at edges. We should be very accurate because it is a prediction model. Prediction model doesn’t always have a hundred percent success rate we can only build them to improve the accuracy at largest possible way.

I have tested as many possible test cases i can that came to my thought and observed whether it is giving accurate results or not.

I have first started testing with the data that have present in the dataset and checked whether the results are accurate and then started checking with the data that is not present in the data set and it gave almost accurate results.

After I finished my testing I asked my team mates to test models so that they may find the drawbacks that I couldn’t because they may have different ideas to find the problems. I have also invited a few of my friends to test the model, aiming to identify any potential drawbacks. By involving multiple individuals in the testing process, we can thoroughly evaluate the system and uncover any existing issues.

**d. What were interesting bugs?**

When it comes bugs other then I have faced the issue with the data I have. The data set I have taken had many changes to be done. The data set was not so perfect to use It consists of many missed values.

The data I have is inconsistent and some columns have data misplaced and few have garbage values. These type of issues may result in the performance of the model and make it inaccurate

To convert the data into perfect data is the biggest challenge to be done and have to perform few operations to do so. Have to remove few rows that have garbage values and to fill the columns with mean or median values that doesn’t effects the accuracy of the model. And few bugs also arrived because of the code errors. These bugs resulted in the inaccuracy of the system. And another challenge is working with categorical data, actually we cannot train the model by using categorical data so, we should convert the categorical data into dummy numerical values and should be very carefull that no data will be missed.

**e. How did you fix them?**

Prioritize the bugs based on their severity, impact, and urgency. This helps allocate resources effectively and address critical issues first

I have deal with the challenges that I have faced about data set by using data preprocessing methods. The data preprocessing steps are getting the dataset, importing libraries, importing datasets, finding missing data, encoding categorical data, splitting dataset into training and test set and feature scaling.

In addition to this project, I have worked on various other projects where I encountered different types of bugs. Many times, the bugs appeared to be elusive, with no compilation errors, but the output was incorrect. In such situations, we resorted to using debuggers, but I preferred a different approach. I would carefully examine the code and strategically place print statements in blocks of code where I suspected the issue might be originating. This allowed me to trace the flow of the program and pinpoint the location of the bug and any logical errors I had made. This method has proven to be effective in resolving those tricky bugs.