

Problem :-

A company works with number of employees, all the works are dependents on the employees. Even if one of the employees resign the job immediately then assigned work will be not finished at the time, so delivery of the project to the clients will be delayed. Company planned to make solution for this, they want to know which employee may resign next. If they know previously, they can arrange alternative to avoid such problem. As an AI Engineer you must give Solution to this.

- A) How will you achieve this in AI?
- B) Find out the 3 -Stage of Problem Identification
- C) Name the project
- D) Create the dummy Dataset.

Name of the project : **Employee Tenure Prediction**

Find out the 3 -Stage of Problem Identification

Stage 1 :-

Domain selection:- Machine learning

I will consider the below important factor

- a) Employee work satisfaction
- b) Employee growth opportunities inside the company.
- c) Employee growth opportunities outside the current company.
- d) Competitor Company salary for the same role
- e) Employee skill set value on Job Market
- f) Employee Manager relationship.
- g) Employee Convenience on his work
- h) Employee Work life balance
- i) Company facilities provided to the employee
- j) Employee Marital status

Interestingly , I do have some numeric data available for comparison .

- a) Employee Current Salary
- b) Employee Tenure
- c) Employee Hike received on last 5 yrs
- d) Employee market salary for his experience
- e) Number of opportunities available for his skills set

This problem is more of sentimental analysis combined with numerical data analysis .

Stage 2:-

Identify the learning selection:-

Since the above mentioned data would be available to the company .

I would consider this case is **supervised learning** .

Note :- Company do have ex ex-employee and current employee data as data set .

Stage 3:-

I would consider the output will be categorial data . So I would be consider this problem . **classification problem** .

Solution : we will consider them as domain , Supervised classification problem category .

Data Set need to have below data as column.

- a) Employee name
- b) Employee ID
- c) Employee sex
- d) Employee Marital status
- e) Employee age
- f) Employee Designation
- g) Employee Current Salary
- h) Employee Tenure
- i) Employee Hike received on last 5 yrs
- j) Employee market salary for his experience
- k) Number of opportunities available for his skills set on current job market
- l) Employee work satisfaction
- m) Employee growth opportunities inside the company.
- n) Employee growth opportunities outside the current company.
- o) Competitor Company salary for the same role
- p) Employee skill set value on Job Market
- q) Employee Manager relationship.
- r) Employee Convenience on his work
- s) Employee Work life balance

- t) Company facilities provided to the employee
- u) Employee Marital status

How will you achieve this in AI?

- a) **Data collection process** : I need to collect above data from HR team and outside source .
- b) **Data preprocessing** : cleaning the data and have the data set in proper way
- c) **Model selection :-** I need to choose a correct suitable model . (ML algorithm)
- d) **Model Training :-** I need to pass or train my model with the data set to the ML algorithm
- e) **Model evaluation and testing :-**
Now , I need to test the model with a given data set

Sample Data Set;-

Note : many columns are available which cant be added in the tabulation which is mentioned above .

Employee ID	Employee Name	Employee Age	Employee Gender	Employee Current Salary	Employee Designation	Employee Tenure	Employee Market Salary for Experience	Resignation Prediction
101	Ravi	35	Male	85,000	Senior Developer	8 years	90,000	No
102	Sita	29	Female	70,000	Project Manager	5 years	75,000	No
103	Michael	42	Male	95,000	Team Lead	10 years	100,000	Yes
104	deepika	37	Female	80,000	Business Analyst	7 years	85,000	No
105	Chenthil	30	Male	65,000	Software Engineer	4 years	70,000	Yes

