**Sub-Functional Area Prediction:**

**Background and Problem Statement** :

Shine.com is an online job discovery portal. Following are some of the attributes the new users are required to fill :

1. **User\_SubFunctional\_Area** : Current sub-functional area of work of a user (Ex. Business intelligence ( User\_SubFunctional\_Area) is a sub-functional Area for Analytics(User\_Functional\_Area))
2. **User\_JobTitle** : The title of job user is pursuing currently.
3. **User\_ProfileTitle** : The title of user’s profile .
4. **User\_Industry** : The Industry user is currently working in .
5. **User\_Experience (Years)** : The experience of user in Years.
6. **User\_Skills** : The skills of the user.

User\_Functional\_Area is an interpreted field from User\_SubFunctional\_Area.

In the attached sample Training dataset (Train\_Data) above mentioned features for Users of a Functional\_ Area (IT - Hardware / Networking / Telecom Engineering) are given. The details of the sub-functional areas of this Functional\_Area are given in the sheet named Lookup\_Subfa.

We need to build a model to predict the sub-functional area of a new user so that the same can be suggested based on features numbered 2 to 6.You need to build a predictive model .

Some points you might want to incorporate :

1. Some users might have filled their Sub-Functional Area wrongly (Their skills might not be matching with their filled Sub-Functional Area). What would you do with their data ? And how would you detect such users ?
2. How would you handle missing data ?

You will be judged on below four parameters:

**Problem Solving Approach** : List down the broad steps you would take for this problem solution, try to be as detailed as possible . Also, specify the broad steps you would take for the following tasks and make codes for the same.

1. Data Cleansing
2. Features Creation/Selection
3. Model Approach
4. Final Prediction and Model Assessment

Please specify the metric chosen by you to validate your model and the prediction accuracy of the same.

You may use R/python as per your convenience.

Please submit your results (Scripts , Prediction Accuracy and Approach followed ) by Monday, 17th July, EOD to vineet.bharti@hindustantimes.com.