

Artificial Intelligence for people engagement

Create an AI solution to achieve these objectives:

- Measure motivation levels of employees
- Flag potential attrition cases based on parameters

Expected solution framework:

Task 1: Feature Selection

- Missing values have already been handled in Sample data. Participants are expected to validate and fix for other data inconsistencies (such as but not limited to outliers, data enrichments etc.)

*** *Data enrichments/transformations as per your judgements - like (**and not limited to**) –*

- *quantitative data conversion to qualitative data if necessary*
- *variable transformation*
- *qualitative variable to be transformation to meaningful factors*

- Sample data has 34 features. Participants are required to identify techniques which will assist in the dimension reduction. Participants must follow the necessary procedures to prepare/treat the variables appropriately before applying a particular dimension reduction technique.

*** *for example when PCA is used for dimension reduction, one has to make sure to normalize all the feature variables and then only apply PCA*

Use multiple techniques together to shortlist model variables.

Some of the most common dimension reduction techniques are:

- *Screening based on correlation/variance*
- *PCA/ICA*
- *Factor Analysis*
- *Cluster analysis*
- *Analysis of variance (ANOVA)*
- *Stepwise/backward/forward variable selections*
- *ML algorithms*

**Participants are encouraged to go beyond above mentioned techniques too*

- Innovation in the data set – Feature engineering - Participants are encouraged to evolve new features by manipulating the existing features.
- Identify the final set of attributes/features that are used for **predicting the attrition probability** in the near future i.e. their attrition behaviour.
- On a separate note perform the following two tasks
 - Find the three most important factors (for e.g. use concept of variable importance) impacting employee **motivation level**. Also, provide appropriate data visualizations (in lines of a dashboard format) indicating the features' impact on **motivation level**.
 - Identify an appropriate basis to categorize/group the employees based on various attributes using suitable unsupervised algorithms to study their characteristics/behaviour pattern. For example, identify groups which are prone to low performance, groups which are less motivated and other groups. Based on this inference/recommendations are to be drawn which can guide the HR to take necessary actions

Task 2: Data visualization and exploratory analysis

Participants are required to perform exploratory analysis to understand their data set better. It is important to draw inferences from these as they aid in building a stable model.

Exploratory analysis may refer to quantitative approach or a qualitative approach (graphical and visualization). These can be performed on a univariate or bivariate or multivariate basis

Some of the common EDA (Exploratory Data Analysis) techniques but not limited to

- Descriptive analysis
- Correlation plots
- Heatmaps
- Barplots for the categorical features
- Kernel Density plots for numeric variables.
- Histograms
- Box plots
- Scatter plots
- Crosstabs

Task 3: Model Proposal

An analytical model framework has to be chosen for sample data. While proposing a model, keep the following points in mind

- It is necessary to split the data into training and testing dataset appropriately.
- The assumption underlying the model framework in general and in particular with the given set has to be illustrated. Participants are expected to clearly indicate the cases under which such assumption may be erroneous / any other limitations associated with the model, leading to instability. A model governance plan to deal with such assumption breakage has to be proposed.
- Unmethodical choosing of models to be avoided. The participant should substantiate why a particular model/ algorithm has been chosen and its appropriateness should be justified either empirically or by a mathematical logic.

Task 4: Model Validation

- Assumption testing – Perform all the statistically test to validate the assumptions in the model. For example, if the model is based out of least squares, you may be validating aspects like no spherical errors etc.
- Performance testing – Participants need to perform extensive performance testing. For example and **not limited to** ,
 - Division of the dataset into train and test sets and analyse/interpret the following:
 - Confusion matrix
 - Concordant/discordant ratios,
 - Sensitivity, Specificity
 - Accuracy, Gini
 - k-fold validation
 - Over/under fitting
- **Benchmark Model:** Build an alternative primitive model/ model with different set of features/ another ML algorithm to benchmark and illustrate using statistical tests on how well your model performs compared to the primitive one

NWM HACKATHON 2019

Task 5: Documentation and Presentation

Participants are required auto-generate a report presenting the results and recommend how to improve employee motivation level, efficiency and retention based on results.

They should also include separate manual covering details related to approach used to solve the problem statement:

Outline would be as follows:

- Executive summary
- Business purpose of the model/ Model design coverage
- Conceptual soundness and Modelling approach/Framework
- Model Development process – Exploratory analysis, variable selection, model description
- Model Validation – results of assumption and performance testing, Benchmark description and comparison
- Benchmark