**Project Summary**

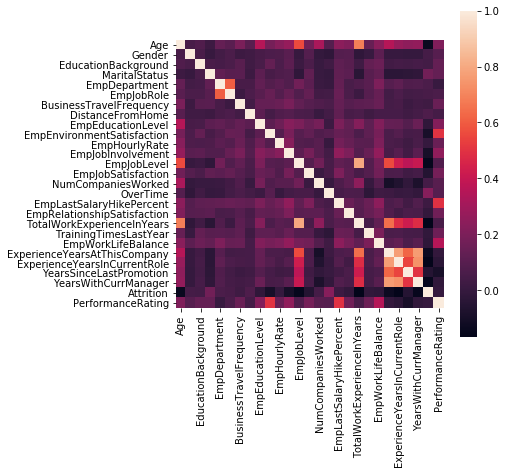
# INX Future Inc Employee Performance:

The Problem statement of the given project clearly specifies that the company INX, is interested in initiating a Data Science project, which analyses the current employee data and find the core underlying causes of this performance issues. And also, the outcomes of the project should help the CEO to take Best actions to improve the performance of the company.

After having a initial look at the Raw Data provided and by looking at the Target variable, the problem can be classified or considered as a Classification Type, as Dependent Variable is Categorical.

Using various Pre-Processing techniques, data is cleaned and different features were selected by having a close look into the correlation between each Feature.

The Below Heat Map plot, which explains the Correlations between features.



# Model Selection:

After Cleaning the data, the data is split into Train and Test sets, which later used to Train Machine Learning models. Since our Project falls under Classification problem type, different Classification models were used to Train and Test. In total, 5 Machine Algorithms, Random Forest, Decision Tree, KNN, SVC, and XGBoost were used. Comparing the Accuracy Score of each algorithm it was clear that Random Forest had the highest accuracy.

The Below table shows ML algorithm and their respective Accuracy Score.

|  |  |
| --- | --- |
| Machine Learning Algorithm | Accuracy Score (in %) |
| Random Forest | 95.12 |
| XGBoost | 94.58 |
| Decision Tree | 93.77 |
| Support Vector Classifier | 75.61 |
| K – Nearest Neighbor | 69.92 |

Random Forest Algorithm, used to Train, Major Features that had influence on the Employee Performances were observed from Model Feature Importance.

# Department wise Performances:

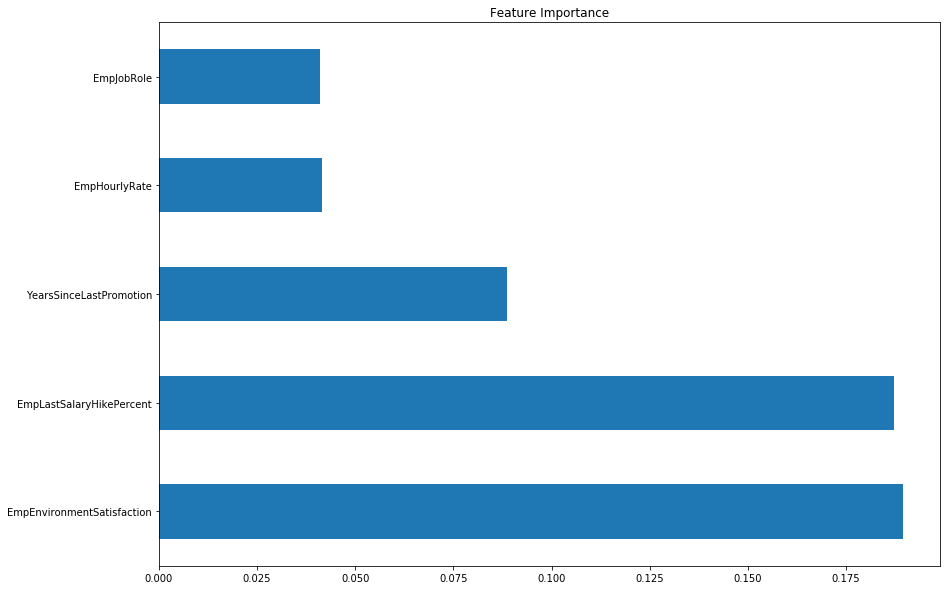
The Below figure shows the Employee Performance across all the Departments available. It can be observed that, Development is clearly a winner, whereas Finance failed perform well while other Departments like Sales, R&D, HR, Data Science performed decently.

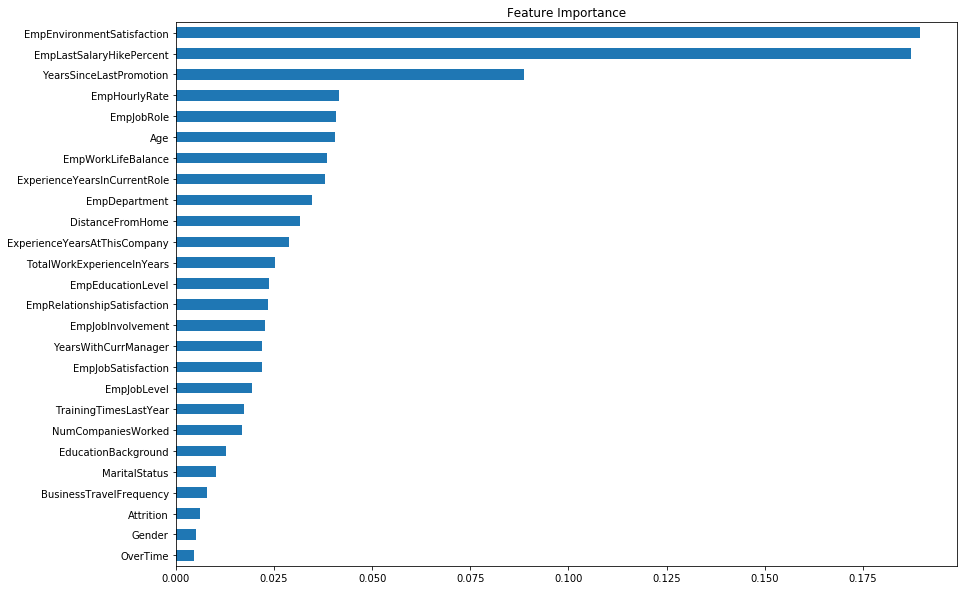


# Feature Importance:

The horizontal Bar graph of Feature Importance, reveals that ‘Employee Environment Satisfaction’ feature had a big influence on **Performance**, whereas ‘Employee Job Role’ feature had the least influence. And it can also be observed that other features like, Employee Hourly Rate, Years Since Last Promotion, Employee Last Salary Hike Percentage had also had a considerable impact on Employees Performance.

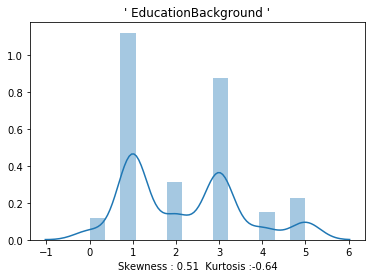
The below Bar graph shows the Feature Importance

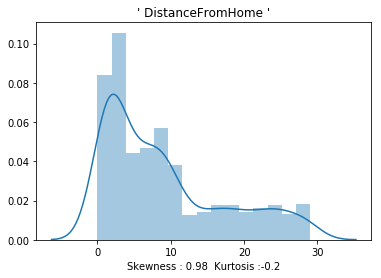


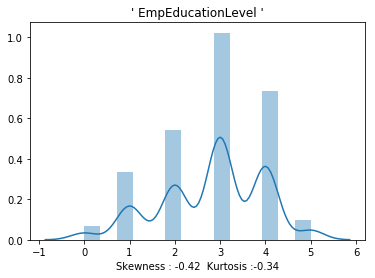


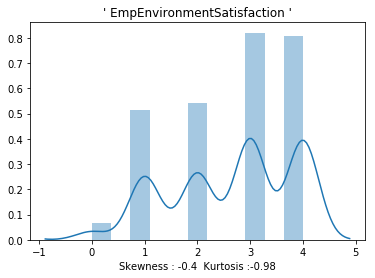
# Skewness v/s Kurtosis:

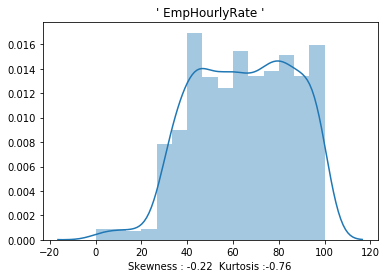
The Below figures shows the Skewness and Kurtosis for different features.

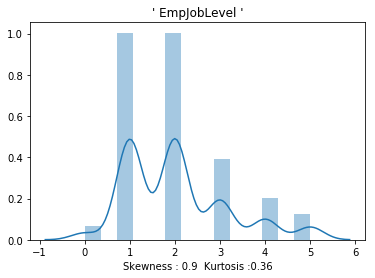


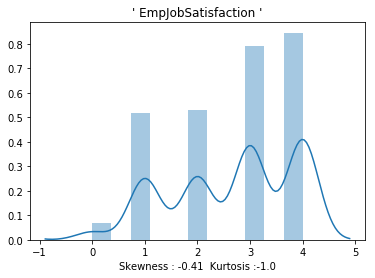


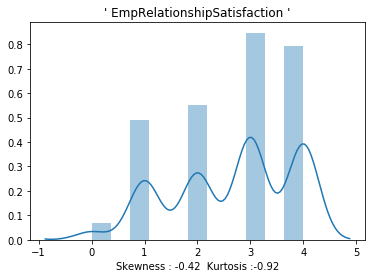


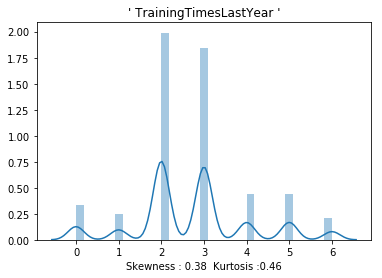


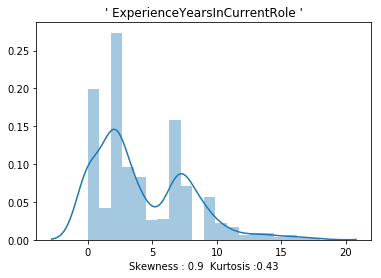


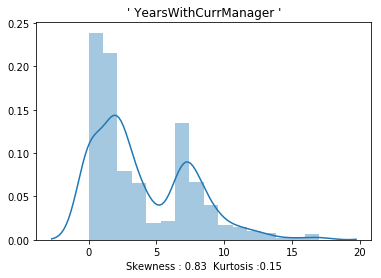












The above figures provide recommendations to improve Employee Performance.

# A Trained Model:

**Creating A Machine Learning Model, involves the following steps.**

* Data Split
* Model Train
* Model Test
* Model Select
* Model Deploy (converting into '.ml')

Finally, a Trained model ‘Employee Rating.ml’ is dumped into Deployment.

Submitted By: Kiran B Kulkarni Assignment ID: E10901-PR2-V18