CASE STUDY - LEAD SCORING

- **Nishant Singh**

Introduction

X Education, an online education company, provides courses for industry professionals through digital platforms. It attracts leads via websites, search engines like Google, and referrals. However, the company faces a low lead conversion rate of around 30%, despite acquiring numerous leads daily. To address this, X Education seeks a predictive model to identify high-potential leads or 'Hot Leads,' helping the sales team focus their efforts effectively. The goal is to improve the lead conversion rate to 80% by assigning a lead score between 0 and 100, indicating the likelihood of conversion.

Business Understanding

X Education aims to enhance its lead conversion rate by identifying high-potential leads from its pool of prospects. The current conversion rate of 30% indicates inefficiencies in targeting the right leads. By building a predictive model to assign lead scores, the company seeks to optimize its sales efforts, prioritize promising leads, and achieve a target conversion rate of 80%, ultimately improving customer acquisition and overall business performance

Problem Statement

The primary objective is to build a predictive model to assign lead scores between 0 and 100, enabling X Education to identify Hot Leads leads likely to convert into paying customers. This will help optimize sales efforts, improve lead conversion rates to the target of 80%, and enhance the efficiency of the sales funnel

If Hot Leads are successfully identified then the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

Approach

1. Data Understanding, Preprocessing and cleaning
2. EDA
3. Modelbuilding using Log Regression and validations

Conclusion

1. Current_Occupation, Last_Activity, Lead_Source, Notable_Activity and Total Time Spent on Website are the most important variables that can be used to predict conversion with high probability
2. Lead_Source_Welingak Website (coefficient 6.14), Current_Occupation_Working Professional (coefficient 3.74) and Lead_Source_Reference (coefficient 3.72) are the top three most important dummy variables which suggests that if the source of the leads are Welingak website or through reference then there are a higher chances of conversion. Also, if the customer is a working professional then again there is a higher chance of conversion
3. Alternately if the Email Bounced then there are the lower chances conversion
4. Following are the results of the final model in place,
    - Accuracy – 80.8%, Sensitivity – 77.8%, Specificity – 82.6%, False positive rate – 17.3%, Precision – 78.9% and Recall – 68.4%
    - ROC curve area – 0.89