WIN PREDICTION
Analytics

Cap Stone Data Science-KPMG

ABSTRACT

In this project, you will analyze and predict the win possibilities of deals/projects for an IT consulting company and see how the possibility of winning a deal is impacted by other variables. This will enable the IT consulting company to manage the effort required to win a deal to meet the growth targets.

Project Outlook

IT firms compete for winning large deals by designing and proposing solutions to their clients. The deal value can reach up to millions of dollars, which leads to highly competitive bidding processes. Even a marginal improvement in the win rate can result into substantial revenue addition for IT firm.

By predicting the probability of winning a deal, the engagement teams can prioritize the pipeline of opportunities to staff the most attractive options first

Objective :-

➢ Predictive Analytics - Build a ML model to predict the probability of win/loss for bidding activities for a potential client.

➢ Prescriptive Analytics – Identify variable/s that are most likely to help in converting an opportunity into a win.

➢ Recommending top 5 combination of SBU Head-Bid Manager.

➢ For every false prediction calculate the loss which the company will face.

Overview of the data:

Column Name	Description
Client Category	Industry in which the client works
Solution Type	The solution group the client requires
Deal Date	The date the opportunity was created
Sector	The sector for which the solution is to be provided
Location	Client location
VP Name	Sr. Manager or VP who is dealing with the client
Manager Name	Manager of the team working on the project
Deal Cost	The initial cost of the deal
Deal Status Code	Final status of the deal(won/lost)

Data Exploration :

- Response variable : Deal Status Code- Won or Loss
- Independent variable : Other 8 variables.
- No of observation : 10061
- Sample data is from the year 2011 to 2019
- Variable type : Data is categorical and most of the variables are in nominal scale except the deal cost (numeric)

Unique values under each category :

Variables	Values
Client Category	41
Solution Type	67
Sector	25
Location	13
VP Name	43
Manager Name	277
Deal Cost	1468
Deal Status Code	2

Data Pre-processing:

Checking for Missing values:

➢ Missing values are considered to be the first barrier in predictive modeling.
➢ The choice of method to impute missing values, largely influences the model's predictive ability.
➢ Deletion is the default method used to impute missing values. But, it not as advisable as it may leads to information loss.
➢ Here it is found that the missing are present in the variables Client category and Deal cost, which may impact the further prediction in the model.

Imputation of missing values:

➢ The missing values in Client category were imputed using dropna in python.

➢ Code:
   df = df.dropna(axis = 0, how ='any')

➢ The missing values in Deal Cost were imputed by using mean method in python.

➢ In this method where the value was 0 in Deal Cost was filled by 768506.6924709034

➢ Code:
   df['Deal Cost']=df['Deal Cost'].replace(0,df['Deal Cost'].mean())

Imputation of duplicate data :-

➢ While searching for duplicate data we found there are 14 duplicate row in data

```
df.duplicated().value_counts()
```
Output= False   9968
           True   14
           dtype: int64

➢

➢ Code :-

```
df=new_df.drop_duplicates()
```

➢ With the help this of this we drop the duplicate data

Influencing factors considered when arriving at the best performing VP:

- No. of years in the company.
- No. of projects handled.
- Value of the projects handled.
- Win Loss Ratio.
- Performance under each independent variables influencing the response variable.

TOP PERFORMERS AMONG VP's

VP name	Project handled from(Year)	Deal countf				
		Total count	Win count	Lost count	Win %	Lost %
Ekta Zutshi	2011	550	344	206	62.5%	37.5%
Sagar Deep Rao	2011	951	346	605	36.4%	63.6%
Longa Bergstrom	2011	908	465	443	51%	49%
Neeraj kumar	2014	469	217	252	46%	54%
Rahul Bujpai	2017	517	203	314	39%	61%%

Thank You