Candidate Solutions Report

Problem Statement: -

Thousands of potential customers visit our website every day for a free horoscope report, some of which actually result in a conversion.

Due to limited human resources, we are unable to reach out to each one of those thousands of potential customers each day. In addition to being infeasible for us, it is probably not necessary either.

To help filter the long list into something manageable by the sales team, we came up with a **baseline model** that prioritizes the customers we reach out to each day. The baseline model was built in a hurry without any serious data analysis and it is just a *static formula* taking as input certain values generated from browsing sessions.

Over the years, we have found that a majority of the potential customers we reach out to do not result in an immediate conversion.

We want to use data and technology to maximize conversions from our contacts each day.

Solution approaches: -

The above problem statement **talks about a baseline model that prioritizes customer based on a static formula**. Now, one can get away with a solution like this that does not involve Machine Learning but it will not be an optimal one. The problem demands a data-driven solution which can be obtained using the following approaches: -

1. Supervised Learning approach

Supervised Learning approach works when we have a target that is mapped to a certain combination of features. Our historic dataset has a target called "**status**" which is what our model can try to learn based on the available features. Since we need to find the probability values associated with each customer ID, we can start building models using algorithms like: - Logistic Regression, Naïve Bayes etc.

2. Unsupervised Learning approach

The approach towards a satisfactory solution does not have to be limited to a supervised algorithm. We can also use unsupervised learning techniques like **Clustering** to cluster our dataset. To apply an unsupervised learning algorithm, we need to have an unlabelled dataset and for our case, we just need to remove the target feature. The Clustering algorithm will group the data based on similarities and differences. Out of the available options, we can go for an approach like "**probabilistic clustering**" to solve this problem.