**Problem Context**

Company XYZ has a list of sales leads that potentially could develop into a business opportunity. To maximize the quality of these leads, XYZ is developing a lead management process, involving system and machine learning algorithms, eventually distributing leads to the sales team for closing. The initial contact lead information (including First Name, Last Name, Mailing Street, Mailing State, Mailing Zip, and Phone) have already been provided. Additional attributes such as “Phone to Name Match" and "Address to Name Match" are generated from matching contacts against XYZ’s customer management system database. The Lead\_Verifiy\_Response file includes a sample of contacts along with each contact’s corresponding attributes(features) and call outcome. Each record/row includes the contact’s name, address, phone number, and the contact’s 10 features along with outcome.

**Input Data**

1. First Name: a contact’s first name
2. Last Name: a contact’s last name
3. Mailing Street: mailing street
4. Mailing State: state
5. Mailing Zip: postal code
6. Phone: a contact’s phone number

**Features**

1. Phone to Name Match: a contact’s name and phone match against CMS database. There are four categories of matching result:
   1. Phone to name Match
   2. No name found for phone number
   3. Phone to name No Match
   4. Invalid phone number
2. Phone Is Valid: phone number is valid, True or False
3. Phone Line Type: Landline, Fixed VOIP, Non-fixed VOIP, ...
4. Phone Prepaid: True or False
5. Phone Is Commercial: True or False
6. Address to Name Match: a contact’s name and address matching against CMS database. There are four categories of matching result:
   1. Address to name match
   2. No name found for address
   3. Either address to name no match
   4. Invalid address
7. Address Is Valid: address is valid, True or False
8. Address Type: e.g. Single unit, Multi unit, ...
9. Address Is Active: True or False
10. Address Is Commercial: True or False

**Outcome(Label)**

1. Call Disposition: ranges from 1 to 6

**Challenge**

Call disposition score of 1 is the primary goal. Score of 2 is the secondarily desired, and 6 is the least desirable. XYZ is trying to predict what feature/attribute or combinations will lead to the highest percentage of score 1? How would you go about solving this problem? What algorithms and methods would you use? What steps would you take? What feature or features produce the best result?