**Dataset Description**

In this dataset, you are provided over a hundred variables describing attributes of life insurance applicants. The task is to predict the "Response" variable for each Id in the test set. "Response" is an ordinal measure of risk that has 8 levels.

**File descriptions**

* train.csv - the training set, contains the Response values.
* test.csv - the test set, you must predict the Response variable for all rows in this file. There are 19766 rows of data, and 127 columns for the test.csv

**Data fields**

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Id | A unique identifier associated with an application. |
| Product\_Info\_1-7 | A set of normalized variables relating to the product applied for |
| Ins\_Age | Normalized age of applicant |
| Ht | Normalized height of applicant |
| Wt | Normalized weight of applicant |
| BMI | Normalized BMI of applicant |
| Employment\_Info\_1-6 | A set of normalized variables relating to the employment history of the applicant. |
| InsuredInfo\_1-6 | A set of normalized variables providing information about the applicant. |
| Insurance\_History\_1-9 | A set of normalized variables relating to the insurance history of the applicant. |
| Family\_Hist\_1-5 | A set of normalized variables relating to the family history of the applicant. |
| Medical\_History\_1-41 | A set of normalized variables relating to the medical history of the applicant. |
| Medical\_Keyword\_1-48 | A set of dummy variables relating to the presence of/absence of a medical keyword being associated with the application. |
| Response | This is the target variable, an ordinal variable relating to the final decision associated with an application |

**The following variables are all categorical (nominal):**

Product\_Info\_1, Product\_Info\_2, Product\_Info\_3, Product\_Info\_5, Product\_Info\_6, Product\_Info\_7, Employment\_Info\_2, Employment\_Info\_3, Employment\_Info\_5, InsuredInfo\_1, InsuredInfo\_2, InsuredInfo\_3, InsuredInfo\_4, InsuredInfo\_5, InsuredInfo\_6, InsuredInfo\_7, Insurance\_History\_1, Insurance\_History\_2, Insurance\_History\_3, Insurance\_History\_4, Insurance\_History\_7, Insurance\_History\_8, Insurance\_History\_9, Family\_Hist\_1, Medical\_History\_2, Medical\_History\_3, Medical\_History\_4, Medical\_History\_5, Medical\_History\_6, Medical\_History\_7, Medical\_History\_8, Medical\_History\_9, Medical\_History\_11, Medical\_History\_12, Medical\_History\_13, Medical\_History\_14, Medical\_History\_16, Medical\_History\_17, Medical\_History\_18, Medical\_History\_19, Medical\_History\_20, Medical\_History\_21, Medical\_History\_22, Medical\_History\_23, Medical\_History\_25, Medical\_History\_26, Medical\_History\_27, Medical\_History\_28, Medical\_History\_29, Medical\_History\_30, Medical\_History\_31, Medical\_History\_33, Medical\_History\_34, Medical\_History\_35, Medical\_History\_36, Medical\_History\_37, Medical\_History\_38, Medical\_History\_39, Medical\_History\_40, Medical\_History\_41

**The following variables are continuous:**

Product\_Info\_4, Ins\_Age, Ht, Wt, BMI, Employment\_Info\_1, Employment\_Info\_4, Employment\_Info\_6, Insurance\_History\_5, Family\_Hist\_2, Family\_Hist\_3, Family\_Hist\_4, Family\_Hist\_5

The following variables are discrete:

Medical\_History\_1, Medical\_History\_10, Medical\_History\_15, Medical\_History\_24, Medical\_History\_32

Medical\_Keyword\_1-48 are dummy variables.