

Thank you for accepting the rules.



\$30,000 • 1,779 teams

Prudential Life Insurance Assessment

Mon 23 Nov 2015

Mon 15 Feb 2016 (34 days to go)

Dashboard

Home



Data



Make a submission



Information



Description

Evaluation

Rules

Prizes

About Prudential

Timeline

Forum



Scripts



New Script




New Notebook

[Competition Details](#) » [Get the Data](#) » [Make a submission](#)

Data Files

| File Name | Available Formats |
|-----------------------|----------------------------------|
| sample_submission.csv | .zip (24.89 kb) |
| test.csv | .zip (819.47 kb) |
| train.csv | .zip (2.42 mb) |

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.

| | |
|----------------|---|
| Leaderboard |  |
| My Team |  |
| My Submissions |  |

Leaderboard

1. Vlad Teodorescu
2. GOOD LUCK
3. Carlos Fernandez
4. raddar
5. library(mlr)
6. river
7. horizon
8. bugfinder
9. wilan
10. xaviercapdepon

967 Scripts

caret_cv
10 Votes / 2 days ago / R

XGBoost with optimized offsets
9 Votes / 10 hours ago / Python

Neural Network Example
19 Votes / 28 days ago / Python

Exploring the Data
25 Votes / 41 days ago / RMarkdown

Features predictability
8 Votes / 22 days ago / RMarkdown

Starter Script
33 Votes / 50 days ago / R

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
- **sample_submission.csv** - a sample submission file in the correct format

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. |

Forum (75 topics)

Mad multi-accounting/leaderboard probing going on
1 hour ago

Help with evaluation score!!!
9 hours ago

Any success with feature engineering or transforming the data?
16 hours ago

xgb_offset0501
20 hours ago

How to do blending
21 hours ago

scikit-learn SVC and LinearSVC to fit model
21 hours ago

1 7 7 9 teams

1 7 7 9 players

2 2 4 7 0 entries

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.

