**Project Overview**

SEND FEEDBACK

**Project Overview**

You are a loan officer at a young and small bank (been in operations for two years) that needs to come up with an efficient solution to classify new customers on whether they can be approved for a loan or not. You'll use a series of classification models to figure out the best model and provide a list of creditworthy customers to your manager.

**How Do I Complete this Project?**

This project uses skills learned throughout the "Classification Models” course. To complete this project:

* Go through the course
* Apply the skills learned in the course to solve the business problem given in the project details section.
* Use our guidelines and rubric to help build your project.
* When you're ready, submit it to us for review using the submission template found in the supporting materials section.

**Skills Required**

In order to complete this project, you must be able to:

* [Cleanup, format, and blend a wide range of data sources](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/ec4b7a7b-1694-4c0c-bfed-224d0733896c)
* Build [predictive classification models](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/7700b31f-1be5-428f-98f0-2dc9b5442690/lessons/3d887284-b0d7-4250-a7d6-626a261d8e26/concepts/af3f276e-4c58-4d3f-b5b6-53d1b55a8d69) using Logistic Regression, Decision Tree, Random Forest, and Boosted Model

NEXT

**Project Details**

SEND FEEDBACK

**The Business Problem**

You work for a small bank and are responsible for determining if customers are creditworthy to give a loan to. Your team typically gets 200 loan applications per week and approves them by hand.

Due to a financial scandal that hit a competitive bank last week, you suddenly have an influx of new people applying for loans for your bank instead of the other bank in your city. All of a sudden you have nearly 500 loan applications to process this week!

Your manager sees this new influx as a great opportunity and wants you to figure out how to process all of these loan applications within one week.

Fortunately for you, you just completed a course in classification modeling and know how to systematically evaluate the creditworthiness of these new loan applicants.

For this project, you will analyze the business problem using the Problem Solving Framework and provide a list of creditworthy customers to your manager in the next two days.

You have the following information to work with:

1. Data on all past applications
2. The list of customers that need to be processed in the next few days

**Steps to Success**

**Step 1: Business and Data Understanding**

Your project should include a description of the key business decisions that need to be made.

**Step 2: Explore and Cleanup the Data**

To properly build the model, and select predictor variables, you need to explore and cleanup your data.

Here are some guidelines to help you clean up the data:

1. Are any of your numerical data fields highly-correlated with each other? The correlation should be at least .70 to be considered “high”.
2. Are there any missing data for each of the data fields? Fields with a lot of missing data should be removed
3. Are there only a few values in a subset of your data field? Does the data field look very uniform (there is only one value for the entire field?). This is called “low variability” and you should remove fields that have low variability. Refer to the "Tips" section to find examples of data fields with low-variability.
4. Your clean data set should have 13 columns where the Average of Age Years should be 36 (rounded up)

**Note:** If you decide to impute any data field, for the sake of consistency in the data cleanup process, impute the data using the median of the entire data field.

**Step 3. Train your Classification Models**

You should choose 70% to create the Estimation set and 30% to create the Validation set. Set the Random Seed to 1 if you're using Alteryx.

Train your dataset using these models:

* Logistic Regression
* Decision Tree
* Forest Model
* Boosted Tree

**Step 4. Writeup**

Compare all of the models’ performance against each other. Decide on the best model and score your new customers.

**Important**: Your manager only cares about how accurate you can identify people who qualify and do not qualify for loans for this problem.

Write a brief report on how you came up with your classification model and write down how many of the new customers would qualify for a loan.

NEXT

### ; [Project:Predicting Default Risk](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436)

## [SEARCH](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/8720c97b-8fc9-498c-ab55-5e898d32deab/lessons/fffc5ef4-ce6a-41e4-b3c5-7a8a36b0ab5d/concepts/ae2a75c9-a9cb-4717-be6d-25c2c55aad12)

## [RESOURCES](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/8720c97b-8fc9-498c-ab55-5e898d32deab/lessons/fffc5ef4-ce6a-41e4-b3c5-7a8a36b0ab5d/concepts/ae2a75c9-a9cb-4717-be6d-25c2c55aad12)

## [CONCEPTS](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/8720c97b-8fc9-498c-ab55-5e898d32deab/lessons/fffc5ef4-ce6a-41e4-b3c5-7a8a36b0ab5d/concepts/ae2a75c9-a9cb-4717-be6d-25c2c55aad12)

1. [1. Project Overview](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/8720c97b-8fc9-498c-ab55-5e898d32deab/lessons/fffc5ef4-ce6a-41e4-b3c5-7a8a36b0ab5d/concepts/4863f554-f6fc-4cd1-9512-292ca84f80be)
2. [2. Project Details](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/8720c97b-8fc9-498c-ab55-5e898d32deab/lessons/fffc5ef4-ce6a-41e4-b3c5-7a8a36b0ab5d/concepts/ac7bf309-3e98-4a2f-ad4e-555905d7cd07)
3. [**3. Supporting Materials**](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/8720c97b-8fc9-498c-ab55-5e898d32deab/lessons/fffc5ef4-ce6a-41e4-b3c5-7a8a36b0ab5d/concepts/ae2a75c9-a9cb-4717-be6d-25c2c55aad12)
4. [4. Tips](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/8720c97b-8fc9-498c-ab55-5e898d32deab/lessons/fffc5ef4-ce6a-41e4-b3c5-7a8a36b0ab5d/concepts/2278b13d-4750-4ac2-8bcc-d5c18eee505f)
5. [5. Verify Answers for Project](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/8720c97b-8fc9-498c-ab55-5e898d32deab/lessons/fffc5ef4-ce6a-41e4-b3c5-7a8a36b0ab5d/concepts/3d8e31d7-94fc-4d84-b1a0-dcabd33178f5)
6. [6. Model Comparison Tool Errors](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/8720c97b-8fc9-498c-ab55-5e898d32deab/lessons/fffc5ef4-ce6a-41e4-b3c5-7a8a36b0ab5d/concepts/e516b125-f3b5-4066-b7ef-bb84e35aca9a)
7. [7. Project: Predicting Default Risk](https://classroom.udacity.com/nanodegrees/nd008t-ent/parts/c44184d8-24f2-4351-8df0-adb09ec4b436/modules/8720c97b-8fc9-498c-ab55-5e898d32deab/lessons/fffc5ef4-ce6a-41e4-b3c5-7a8a36b0ab5d/project)

* Mentor Help

Ask a mentor on our Q&A platform

* Peer Chat

Chat with peers and alumni

# Supporting Materials

SEND FEEDBACK

### Review

Use the [project rubric](https://review.udacity.com/#!/rubrics/265/view) to review your project. If you are happy with your submission, then you're ready to submit your project. If you see room for improvement, keep working to improve your project.

### Submission Template

Use the submission template at the bottom of this section to submit your project. After filling it out, save it as a PDF and submit the PDF in the next section. You may also include your Alteryx workflow if you'd like. If your submission does not meet specifications, having the workflow may help the review identify mistakes.

### Workflow Template

Use the Alteryx workflow template to help quickly get started with the project. For the sake of the reviews consistency, you must use this Alteryx workflow.

#### Data

credit-data-training.xlsx - This file contains all credit approvals from your past loan applicants the bank has ever completed.

customers-to-score.xlsx - This is the new set of customers that you need to score on the classification model you will create.

### Non Alteryx Software

For students using software other than Alteryx to create the classification models, make sure to format the data as follows and do not sort the data before you run it through your classification models. This allows you to get consistent results reviewers expect.

| **Variable** | **Data Type** |
| --- | --- |
| Credit-Application-Result | String |
| Account-Balance | String |
| Duration-of-Credit-Month | Double |
| Payment-Status-of-Previous-Credit | String |
| Purpose | String |
| Credit-Amount | Double |
| Value-Savings-Stocks | String |
| Length-of-current-employment | String |
| Instalment-per-cent | Double |
| Guarantors | String |
| Duration-in-Current-address | Double |
| Most-valuable-available-asset | Double |
| Age-years | Double |
| Concurrent-Credits | String |
| Type-of-apartment | Double |
| No-of-Credits-at-this-Bank | String |
| Occupation | Double |
| No-of-dependents | Double |
| Telephone | Double |
| Foreign-Worker | Double |

### Project Checklist

This is here to make sure you know the concepts necessary to complete the project and that you know where to go find answers if you need help. If you do not feel confident in the topics addressed please review the course material or reach out on Knowledge or Study Groups!

Task List

* 

Showed which predictor variables are significant or the most important.

* 

Showed the p-values or variable importance charts for all of the models.

* 

Provided justification for why I imputed or removed these fields from the data.

* 

I checked the rubric <https://review.udacity.com/#!/rubrics/265/view>

* 

I have gone to Knowledge or Study Groups before submitting

We have received some feedback around the legitimacy of excluding the Telephone variable since it potentially could be useful to collect debts so could make you more creditworthy. We think this is fair but for the sake of this project please exclude Telephone from your data with the reasoning that there is no logical reason for including the variable.

* Forums: <https://discussions.udacity.com/c/nd008-bizand-classification-models/band-project-4>
* Study Groups

If you run into errors in Alteryx or unexpected results from a tool we have a guide to help you figure out what is going on.

**Alteryx Debugging Guide:**

Click and open the Resources Tab in the left most panel of your classroom to find the Alteryx Debugging Guide or here's a [Google Doc](https://docs.google.com/document/d/1ec7SiDCMAZAMjbPvL3aegPoL1PZznHarsxEZ4bjptdI/edit?usp=sharing), please note if you are accessing Udacity from a corporate network you may not be able to view the Google Doc, please use the Resources tab instead.

**To Download Files below please right click on the link and select "Save Link As"**

#### Supporting Materials

* [Credit Data Training](https://video.udacity-data.com/topher/2016/October/57fd7d05_credit-data-training/credit-data-training.xlsx)
* [Customers to Score](https://video.udacity-data.com/topher/2016/October/57fd7d13_customers-to-score/customers-to-score.xlsx)
* [Alteryx Workflow Template](https://video.udacity-data.com/topher/2017/February/58990dbf_project-template/project-template.zip)
* [Submission Template](https://video.udacity-data.com/topher/2018/February/5a8117ad_submissiontemplate-3-1/submissiontemplate-3-1.docx)
* [All Project Files](https://video.udacity-data.com/topher/2018/February/5a8117e5_creditworthiness/creditworthiness.zip)

NEXT

;