

Capstone Project Plan:

Validation of machine learning methods for classification of life insurance risk

Introduction: In the life insurance industry, underwriting can be a time-consuming task which can involve a degree of subjectivity across customers. Our goal is to determine whether machine learning model(s) can be used to classify potential life insurance customers into specific risk categories accurately and efficiently thereby eliminating any subjectivity in the process, reducing time spent, and ultimately saving on costs. The best model will be selected based on its accuracy and time taken for prediction.

Project Management Approach: The team consists of four members: Brian, Chris, Luke N., and Daoud. The work will be shared across the group equally (i.e., every step is agreed upon by the group). This will be done using shared workspace platforms (like shared word docs, data bricks, etc.). Most communication will be handled over MS Teams when necessary.

Project Scope:

In-Scope: What is in-scope for this project is determining whether using machine learning models is an efficient way to classify customers into life insurance risk categories (one through eight). To that end the best models will be found and a recommendation will be given for the best model assuming it is more accurate than a cut-off value (TBD).

Out-of-Scope: What's out-of-scope for this project, but would be great for further study, include things like:

1. What feature(s) of their combinations lead to specific risk category placement?
2. Is there a way to quantify how similar customers are to one another?
3. Are results consistent across geographic and temporal settings?
4. Etc.

Milestones:

| Name | Description | Due Date |
|-------------------------|---|---------------------------------|
| Project Management Plan | A plan detailing the purpose of this project, who is involved, what work will be done, and the timeline over which it will be performed | Friday, Oct. 29 th |
| Exploratory Questions | The data to be explored will be gathered and based off this data, 5 to 10 questions to explore will be created and put into a pdf file. | Friday, Oct. 29 th |
| ETL Report | Extract-Transform-Load: Describe data, how it was extracted, from where, | Wednesday, Nov. 3 rd |

| | | |
|------------------------|---|---|
| | transformation steps, how and where it was loaded in. | |
| Napkin Drawing (V) | Draft of visuals based off of the cleaned data. Feedback. | Draft: Tuesday, Nov 2 nd Feedback: Wednesday, Nov 3 rd |
| Napkin Drawing (DB) | Draft of dashboard based off the report, including visuals, summary/about. Feedback. | Draft: Tuesday, Nov 2 nd Feedback: Wednesday, Nov 3 rd |
| PowerBI Dashboard | A dashboard made in PowerBI based off the napkin drawing in the step above. | Monday, Nov. 8 th |
| Project Summary/Report | Intro and purpose, process of obtaining and using data, analysis, and conclusion. | Wednesday, Nov. 10 th |
| Presentation Creation | Creation of 10 min long presentation. Will be done in PowerPoint. Details the introduction, data set, data processing and methodology, and results and recommendations. | Thursday, Nov. 11 th |
| Project Presentation | Will consist of both a dry run and actual | Dry run: Thursday, Nov. 11 th Actual: Friday, Nov. 12 th |
| GitHub | Will upload all aspects of the project into GitHub and will create a readme.md for the project and repository. | Thursday, Nov. 11 th |

Real Time schedule: <https://trello.com/b/cdFC6CYa/capstone-group3>