**Data Science Institute of Australia**

Capstone Project

Due Date: 29/07/2019

|  |  |
| --- | --- |
| **Epic** | Auto Insurance Policy Classification |
| **Project Status** | Development |
| **Client** | AHI, Angus Macdougal (IT Manager) |
| **Project Owner** | Sanjee |
| **Developers** | Sanjee |
| **Stakeholders** | Underwriters, Insurers, Policyholders, Brokers |
| **Design Complete** | 31/07/2019 |
| **Code Complete** | 31/08/2019 |
| **Testing Complete** | 30/09/2019 |
| **Release Date** | 31/10/2019 |

**Background**

Policy holders less likely to make a claim are more profitable for the business. These policy holders are referred to as passive policy holders. Currently at the moment there is no mechanism in place to predict the risk factor of a new business policy. This presents an opportunity to improve profitability.

The desired outcome of the project is to predict the risk factor of a new business policy. Premiums can then be adjusted dynamically to attract more passive policies.

**Objectives**

|  |  |
| --- | --- |
| **Objectives** | **Metric** |
| Predict likelihood of a claim for a policy | ‘Claim’, ‘No Claim’ |
| Accuracy of the prediction | Percentage |
| Predict Claim type of New Business policies | ‘Passive’, ‘Attrition’ |
| Accuracy of the prediction | Percentage |

**Assumptions**

* It is assumed that when premium for a predicted attrition policy is increased client will not swing to a competitor.

**User Stories & Requirements**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ref #** | **Name** | **User Story** | **Functional Requirements** | **Notes** |
| 1 | Premium pricing | As an underwriter I want to know after predicting the type of policy by how much I should adjust the premium | Pricing calculator |  |
| 2 | Website | As a underwriter I want an easy way to input parameters | Parameter input GUI |  |
|  |  |  |  |  |

**Open Questions**

|  |  |  |
| --- | --- | --- |
| **Question** | **Answer** | **Date Question Answered** |
| Costs associated with implementation/deployment of both models | Will have to estimated |  |
|  |  |  |

**Out of Scope**

* Get data for the both the models from the same source.
* After model deployment creating monitoring dashboards to evaluate model performance on live data.
* Refresh training data by feeding live data.

**Risks**

|  |  |  |  |
| --- | --- | --- | --- |
| **Risks** | **Likelihood** | **Impact** | **Migration** |
| Misclassification of a Policy | Medium | High | Double check the probability on both the models. Only proceed if they both show a high probability. |
| XGBoost might overfit on the training model | Medium | Medium | Should balance accuracy by evaluating with the test set. |

**User Interaction and Design**

****

**Tasks**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tasks** | **Name** | **Due Date** | **Time Spent/Estimated** |
| Data Integration | DBA | 31/07/2019 | At least 1 month estimated |
| Data Transformation | Sanjee, DBA | 31/08/2019 | 20 hours |
| Feature Selection | Sanjee | 15/09/2019 | 15 hours |
| Model Selection and Evaluation | Sanjee | 30/09/2019 | 15 hours |
| Model Deployment | Out of Scope | 31/10/2019 | 1-2 weeks |
| Website and UX Development | Out of Scope | 15/11/2019 | 2-4 weeks |
| Production Model Monitoring and Evaluation | Out of Scope | Out of Scope | Will have to re-evaluate at monthly intervals |

**Results and Conclusions**

* Overall the model could provide the underwriters a user-friendly tool for risk classification.
* This would in turn help make better pricing decisions and reduce the liability of the Insurers.
* Improvement in work flow efficiency, added revenue through premiums and reduced liability will be welcome for an Underwriting Agency or Insurance Company.