**OPIM 5604-B13: Predictive Modeling**

**Project Proposal; PROGRAM: MSBAPM**

**Topic: Claim re-negotiation prediction**

**Team: Predictive modelling Team ISSSR**

**Roshan Zubair, Sandeep Matta, Suyash Tiwari, Ishrath Sharffudin, Sneha Prabhu**

**Data Source: http://www.scan-support.com/help/sample-data-sets**

**Problem Statement/Objective: To predict if a customer is going to enter re-negotiation with the Insurance company over the approved claim amount.**

**Summary Statistics of the dataset:**

|  |  |  |
| --- | --- | --- |
| **Description** | **Metric** | **Comments** |

Unit of analysis (Observation unit) Claim level

Number of Records 9134 No NAs

Target Variable Response Yes or no

Number of Predictors available 23

Number of Predictors to be used 10 Boolean vars?

**Plan of Action: The claims dataset will be modelled to predict if a customer is going to try to re-negotiate with the client based on the claim amount. For this the modelling process we will employ will follow this path**

1. **Exploratory data analysis – Try to understand how predictors behave, both in a univariate and multivariate sense. This will be done by building Scatterplots, Box plots, Correlation matrix, Contingency tables, etc**
2. **Variable transformation and selection – Try to boost predictor power by creating transformed variables or indicator variables and select the right variables against the target variable (Multivariate tables or mixed/forward selection techniques)**
3. **Model building – The selected variables will be used to build a model, after many iterations, to attain maximum accuracy, while checking for multicollinearity using VIF metric. We will adopt a Train-validation split for the initial model step**
4. **Model evaluation – The finalized model will be checked for accuracy by K fold cross validation. Model will be used as a benchmark to check if machine learning models have better accuracy.**
5. **Story boarding – Using the model and Exploratory data analysis we will try to outline what drives a customer to re-negotiate with client. We will try to come up with additional insights that can be a value add to the organization.**

**Tools/Technology:**

**Tools to be used – R, Python(optional), Jmp**

**OPIM 5604-B13: Predictive Modeling**

**Alternate Project Proposal; PROGRAM: MSBAPM**

**Topic: Tele-com churn**

**Team: Predictive modelling Team ISSSR**

**Roshan Zubair, Sandeep Matta, Suyash Tiwari, Ishrath Sharffudin, Sneha Prabhu**

**Data Source: http://www.scan-support.com/help/sample-data-sets**

**Problem Statement/Objective: To predict the customer**

**Summary Statistics of the dataset:**

|  |  |  |
| --- | --- | --- |
| **Description** | **Metric** | **Comments** |

Unit of analysis (Observation unit) Claim level

Number of Records 7033 No NAs

Target Variable Response Yes or no

Number of Predictors available 20

Number of Predictors to be used 10 Boolean vars?

**Plan of Action: The claims dataset will be modelled to predict if a customer will churn out based on his cellphone plan and usage**

1. **Exploratory data analysis – Try to understand how predictors behave, both in a univariate and multivariate sense. This will be done by building Scatterplots, Box plots, Correlation matrix, Contingency tables, etc**
2. **Variable transformation and selection – Try to boost predictor power by creating transformed variables or indicator variables and select the right variables against the target variable (Multivariate tables or mixed/forward selection techniques)**
3. **Model building – The selected variables will be used to build a model, after many iterations, to attain maximum accuracy, while checking for multicollinearity using VIF metric. We will adopt a Train-validation split for the initial model step**
4. **Model evaluation – The finalized model will be checked for accuracy by K fold cross validation. Model will be used as a benchmark to check if machine learning models have better accuracy.**
5. **Story boarding – Using the model and Exploratory data analysis we will try to outline what drives a customer to re-negotiate with client. We will try to come up with additional insights that can be a value add to the organization.**

**Tools/Technology:**

**Tools to be used – R, Python(optional), Jmp**