

# Allstate\_Claims

February 17, 2018

## 1 Machine Learning Engineer Nanodegree

### 1.1 Allstate Insurance Capstone Proposal

Jose Manuel Garcia

December 31st, 2050

Allstate Claim Severity Predictive Model

#### 1.1.1 Domain Background

Allstate is one of the largest personal insurance companies in the United States and protect millions of people from potential damages. When car accidents happen or when other unexpected disasters occur it is a huge relief knowing you are covered by some type of insurance. Recently my girlfriend was involved in a car accident that almost totaled her car. Although she escaped with no bodily injuries her car had thousands of dollars in damages. Luckily, she has full coverage insurance and was able to get all repairs paid for and a rental car without paying anything out of pocket. This event peaked my interest in the insurance business which led me to this project. Although for this particular instance the insurance claim process was easy and straightforward it is not always the case. This is why Allstate is currently developing automated methods of predicting the cost, and hence severity, of claims. This project will provide insight into better ways to predict claims severity.

#### 1.1.2 Problem Statement

Allstate is attempting to improve its customer experience by automating methods of predicting the severity of the claim. For this project Allstate provides all data points necessary for predicting the severity of the claim.

#### 1.1.3 Datasets and Inputs

Allstate provides a train and test data set for its Claim Severity competition. Both the training and testing data sets from with many parameters that are used to determine the severity of the claim and the training set also has the loss value of the claim.

#### 1.1.4 Solution Statement

The data set will be evaluated thoroughly in order to get an idea of which supervised learning technique would be best for this particular case. There are many data points that are provided

so the first thing that will need to be done is to figure out which data points are relevant to the severity of the claim. The next step would then be to tune the parameters of the supervised learner to get the best possible solutions that does not over fit the data.

### **1.1.5 Benchmark Model**

For this project Allstate provides a training set that has the actual value of the claim severity.

### **1.1.6 Evaluation Metrics**

The evaluation metric that will be used for this project is the R2 score. The R2 score is the proportion of variance in the dependent variable that is predictable from the independent variable. An R2 score of 0 means that the dependent variable cannot be predicted from the independent variable. An R2 score of 1 means that the dependent variable can be predicted from the independent variable. The ideal model should have and R2 score of 1 but the objective will be to get it as close to 1 as possible.

### **1.1.7 Project Design**

A big part of machine learning is acquiring the data to build you model and making sure it is clean and reliable. For this project Allstate has already done most of this work by providing all of the data for this Kaggle competition. The next step will be to process the data so that it be used by classifiers in scikit learn. The classifiers in scikit learn require data to be in number format so all letters will need to be converted in to numbers or hot keys. Some of the data is number format and will need to be normalized so that it can be processed and visualized better. After the data is processed a classifier will be used to train the model. Once a classifier is chosen the parameters of the classifier will be tuned until we get the best possible model.

<https://www.kaggle.com/c/allstate-claims-severity>