

# Project 1 Report

Dreycen Foiles, Olek Yaldas

September 29, 2022

## 1 Introduction

We report on our efforts to use different regression techniques to predict the price of a house given various different parameters. The two techniques that we present in this report is the Least Absolute Shrinkage and Selection Operator (LASSO) and the Random Forest regression algorithm.

## 2 Data Preprocessing

For this project, we are given a data set made up numerous data points. These include basement area, zoning type, heating type, year built, etc. Our goal is to use these values to predict the price of the corresponding house. Before we can build a model, we must convert non-numerical data into a form that can be used with our regression techniques. To do this, we use a very simple conversion where every unique string in a column corresponds to a number. This is accomplished automatically using the Pandas function `factorize`. In addition to the numeric conversion, we also remove some columns that are