Stat 220 Lab 10: Developing Interpretable Models for Construction Project Estimations

1   Introduction

You are now part of the Advanced Construction Analytics Division (ACAD) at Global Building Insights (GBI), a pioneering consultancy in the construction sector. ACAD is renowned for merging statistical rigor with practical utility in its analytical models. Our clients, primarily Construction Estimators, rely heavily on these models to predict project timelines accurately. However, they emphasize the need for these models to be interpretable, allowing for clear understanding and communication of results.

In response to this need, your task is to develop a model that not only predicts the completion time of construction projects with high accuracy but also maintains a level of simplicity and interpretability. This lab focuses on striking the right balance between these two often competing aspects of model building.

Data Acquisition and Initial Exploration

Begin with a thorough examination of the dataset.

1. Access the dataset at `https://richardson.byu.edu/220/construction.csv`

2. Conduct an exploratory analysis to gain insights into the dataset?s characteristics.

Balancing Accuracy with Interpretability

Explore the dual objectives of accuracy and interpretability:

1. Develop an initial linear regression model, considering all available predictors.

2. Evaluate the model's accuracy using metrics like MSE and R^2.

3. Refine your model by selecting predictors and considering potential transformations or interactions, focusing on maintaining interpretability.

4. Discuss the trade-offs made between model complexity and its interpretability.

Model Application and Interpretation

Applying the model to new data:

1. Retrieve a new dataset for prediction from `https://richardson.byu.edu/220/construction_new.csv`.

2. Apply your model to predict the timeline for the new project.

3. Provide a confidence interval for your prediction and interpret it in a manner understandable to Construction Estimators.

Deliverables

Your work will culminate in two key deliverables:

1. A script or notebook containing all analyses and modeling steps.

2. An insightful report for ACAD, focusing on:

   - How the final model achieves a balance between accuracy and interpretability.
   - Justification of model choices with respect to this balance.
   - Insights and actionable recommendations derived from the model.