Introduction

Title: Advanced Analytics for Predicting Business Performance Scores

Objective: To utilize Generalized Additive Models (GAM) and Decision Trees to forecast and analyze the performance ratings of online firms. This study aims to uncover the fundamental patterns and correlations inside the data that have a substantial impact on business evaluations.

Table of Contents

1. Introduction and Background

The modern digital marketplace is a dynamic setting where online enterprises constantly adjust to changing consumer demands and competition landscapes. The capacity to precisely forecast corporate performance through analytical insights confers a significant edge. This paper seeks to utilize sophisticated analytics approaches to forecast the performance scores of online firms, using a comprehensive dataset to examine the effectiveness of several prediction models.

2. Data Understanding

The dataset consists of various attributes such as city, state, score, review count, gender, CEO school category, CEO graduation year, and other relevant business metrics. Data preprocessing involved handling missing values, outlier detection, and data normalization to prepare the dataset for further analysis.

3. Methodology

The analysis follows the CRISP-DM framework:

1. **Business Understanding**
2. **Data Understanding**
3. **Data Preparation**
4. **Modeling**
5. **Evaluation**
6. **Deployment**

4. Feature Engineering

Feature engineering involved transforming categorical variables into dummy variables, normalizing numerical variables, and imputing missing values using median and mode imputation methods for numerical and categorical variables, respectively.

5. Exploratory Data Analysis (EDA)

EDA included visualizing the distribution of scores across different states, cities, and other variables, as well as identifying patterns and correlations within the dataset.

6. Model Implementation

The main modeling techniques used in this analysis are:

1. **Lasso Regression**
2. **Generalized Additive Models (GAM)**
3. **Decision Trees**