Introduction

Title: Advanced Analytics for Predicting Issue Consequences

Objective: To investigate a dataset to forecast the “Issue Consequence” variable using the CRISP-DM framework, including Lasso Regression, Logistic Regression, and K-Nearest Neighbors (KNN) models.

Table of Contents

1. Introduction and Background

The aim of this analysis is to use various predictive modeling techniques to forecast issue consequences. The CRISP-DM framework guides the methodology, ensuring a structured approach to data mining.

2. Data Understanding

The dataset consists of 566,760 observations and 77 variables, including numeric, text, and date variables.

3. Methodology

CRISP-DM Process:

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

4. Feature Engineering

Feature engineering involves transforming data to enhance model performance.

5. Exploratory Data Analysis (EDA)

Various visualizations and statistical summaries are used to understand the data distribution and relationships.

6. Model Implementation

The main modeling techniques used in this analysis are:

1. **Lasso Regression**

2. **Logistic Regression**
3. **K-Nearest Neighbors (KNN)**