Data Mining
Final Group Presentation
(US Accidents Dataset)

Group 7

Gregg Legarda

Daniel Frey

Siwei Wang

Introduction

- Our dataset
- Pre-processing and Data Analysis
- Data Visualization
- ML Algorithms
- Demonstration

Description of Dataset

- Dataset

- Target choosing

- Features variables

Preprocessing

- Categorical vs. Numeric Data
- Missing data (mean, mode, "0")
- Scaling (orders of magnitude)
- PCA (high dimensionality)
- Gridsearch CV

Models

- Training Split, n samples
- Algorithms
- Accuracy
- Prediction

Application

- Summary of the Statistics
- Graphs (Histograms, Correlation Matrix)
- Geo Maps (Plotted data points)
- Clustered Geo Maps
- Prediction of Severity of Accident

Added Features

- Dark mode
- Status Bar (Model and Map)
- Sync Map to status bar
- Pop ups for error handling

Demo

- Video
- Live