SafeRoad

CS 4624
Virginia Tech, Blacksburg, VA
4/29/16

Matthew Longobardi Paradiso
Matthew Morrison
Julio Suriano Siu

Client: Xuan Zhang

Introduction

SafeRoad’s goal:

- Analyze common complaints
- Predict car recalls

NHTSA: National Highway Transportation Safety Administration

Use machine learning to predict common complaints

Evaluate and improve the results

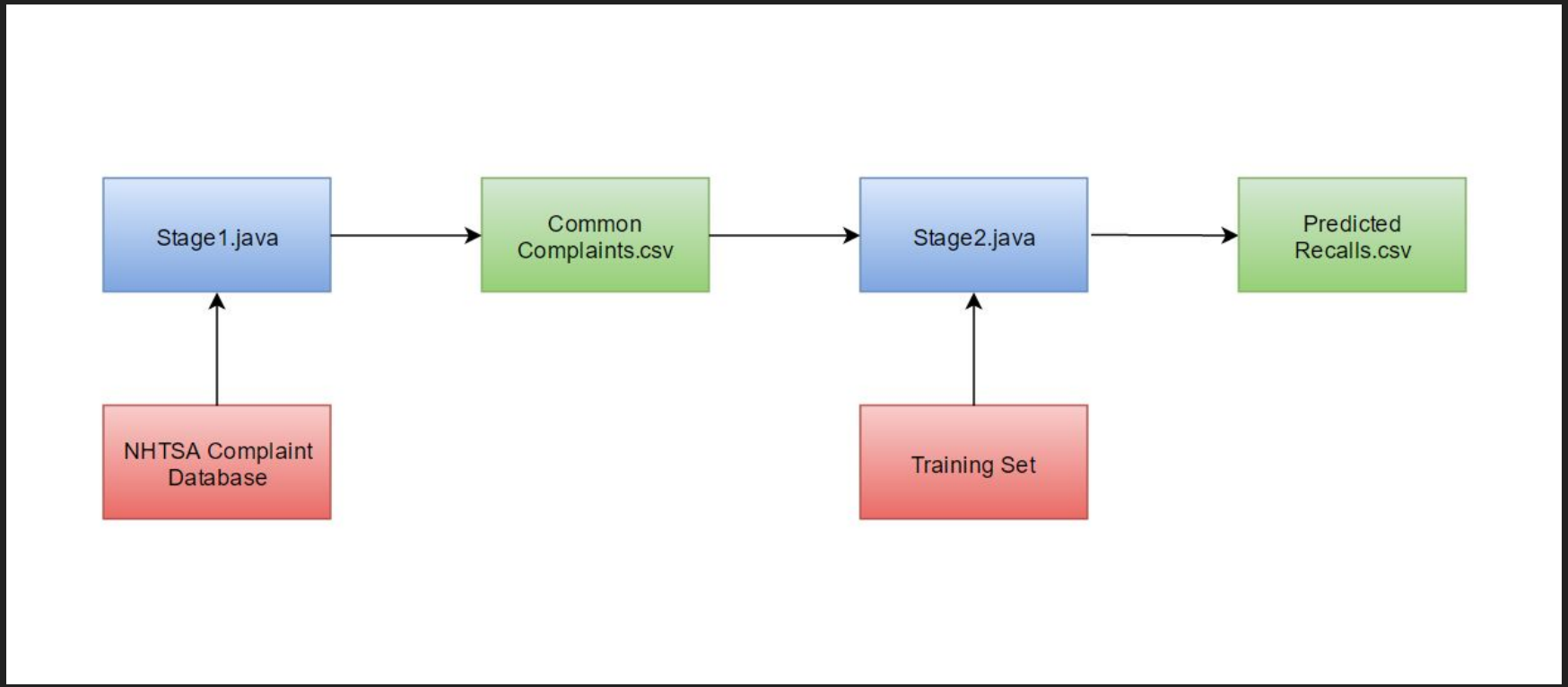
Data and Tools

NHTSA databases:
- Complaint database
- Recall database

Java DataBase Connection (JDBC) - access SQL database through Java

Java Machine Learning (JML) - provides functionality for classifying complaints

Architecture of the Program



Raw complaint data:

500001,10089761,HONDA (AMERICAN HONDA MOTOR CO.),HONDA,ELEMENT,2003,N,20040825,N,0,0,
VISIBILITY:WINDSHIELD,FALLS CHURCH,VA,5J6YH28503L,20040825,20040825,,1,"WHILE DRIVING A
ROCK FROM THE STREET HIT THE WINDSHIELD, CAUSING A CRACK THREE TO FOUR INCHES IN DIAMETER.
CONSUMER DROVE THE VEHICLE TO THE DEALER FOR INSPECTION, AND MECHANIC DETERMINED THAT THE
ROCK WAS THE CAUSE OF THE PROBLEM. CONSUMER INFORMED THE MECHANIC THAT THIS PROBLEM
OCCURRED LESS THAN A YEAR AGO, AND   PROBLEM RECURRED.*AK",EVOQ,N,,Y,Y,Y,4,4WD,FI,DS,AUTO,
55,,,,,,,,,,,,,V,

Processed complaint data:

2003, HONDA, ELEMENT, VISIBILITY, N, N, 0, 0, "WHILE DRIVING A ROCK FROM THE STREET HIT THE
WINDSHIELD, CAUSING A CRACK THREE TO FOUR INCHES IN DIAMETER. CONSUMER DROVE THE VEHICLE
TO THE DEALER FOR INSPECTION, AND MECHANIC DETERMINED THAT THE ROCK WAS THE CAUSE OF THE
PROBLEM. CONSUMER INFORMED THE MECHANIC THAT THIS PROBLEM OCCURRED LESS THAN A YEAR AGO,
AND   PROBLEM RECURRED.*AK", U

Classifier

Use machine learning classifiers to classify complaints

JML classifiers work with numbers

Use a HashMap to map words to numbers

Use 1’s and 0’s for boolean values

```
2004,Honda,Element,Service Brakes,N,N,0,0,”Vibration when braking”,N
```

```
2004,7,201,103,0,0,0,0,”Vibration when braking”,N
```

HashMap

1 FORD	11 PONTIAC	21 CADILLAC	31 RAM
2 CHEVROLET	12 VOLKSWAGEN	22 LEXUS	32 OLDSMOBILE
3 TOYOTA	13 BMW	23 VOLVO	33 LAND ROVER
4 DODGE	14 SATURN	24 ACURA	34 SAAB
5 JEEP	15 KIA	25 SUZUKI	35 SCION
6 NISSAN	16 MERCEDES BENZ	26 MINI	36 ISUZU
7 HONDA	17 MAZDA	27 MITSUBISHI	37 YAMAHA
8 CHRYSLER	18 BUICK	28 INFINITI	38 JAGUAR
9 HYUNDAI	19 MERCURY	29 LINCOLN	39 HUMMER
10 GMC	20 SUBARU	30 AUDI	40 DAEWOO

Prototype and Testing - F-Score

F-test score is a measure of accuracy for binary classification

F = 2 * (P * R) / (P + R)

P = correct recall classifications / all recall classifications

R = correct recall classifications / expected number of recall classifications

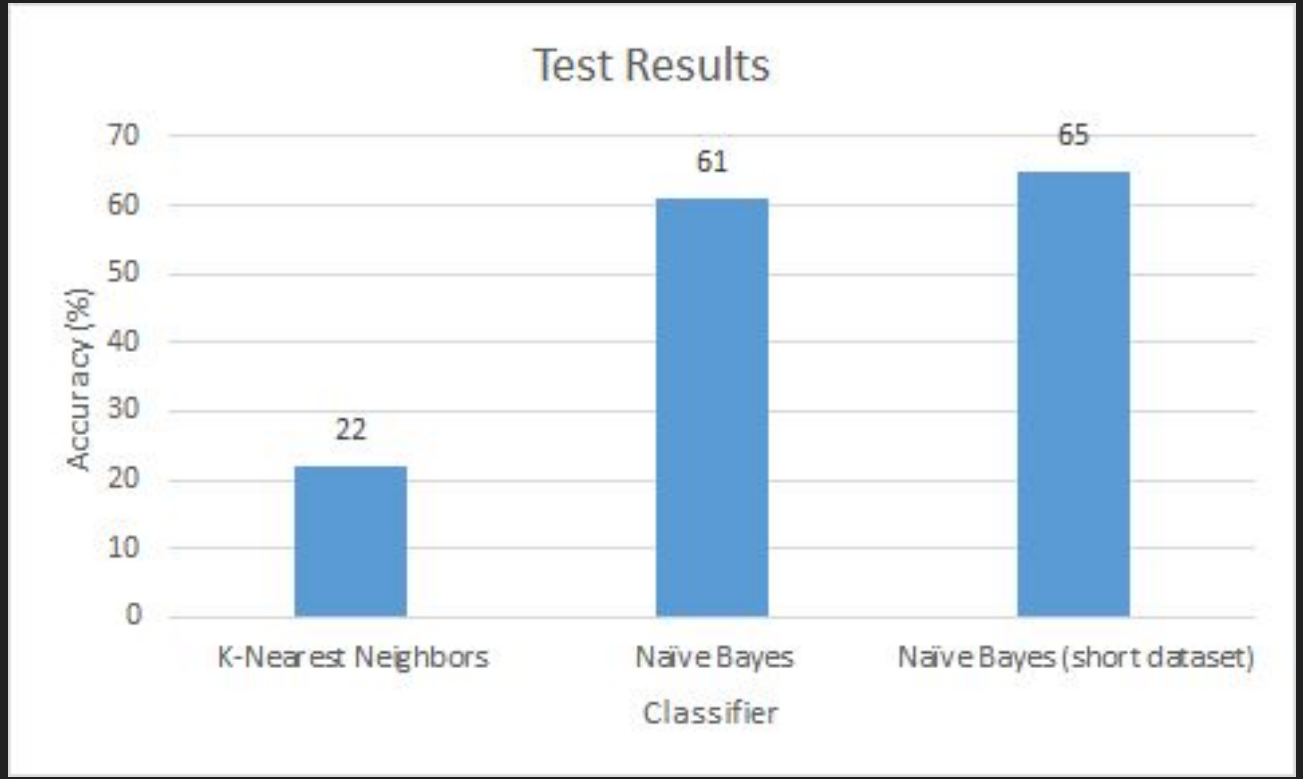
Prototype and Testing

Tested prototype with training set containing 350 records

Performed 3 different tests:

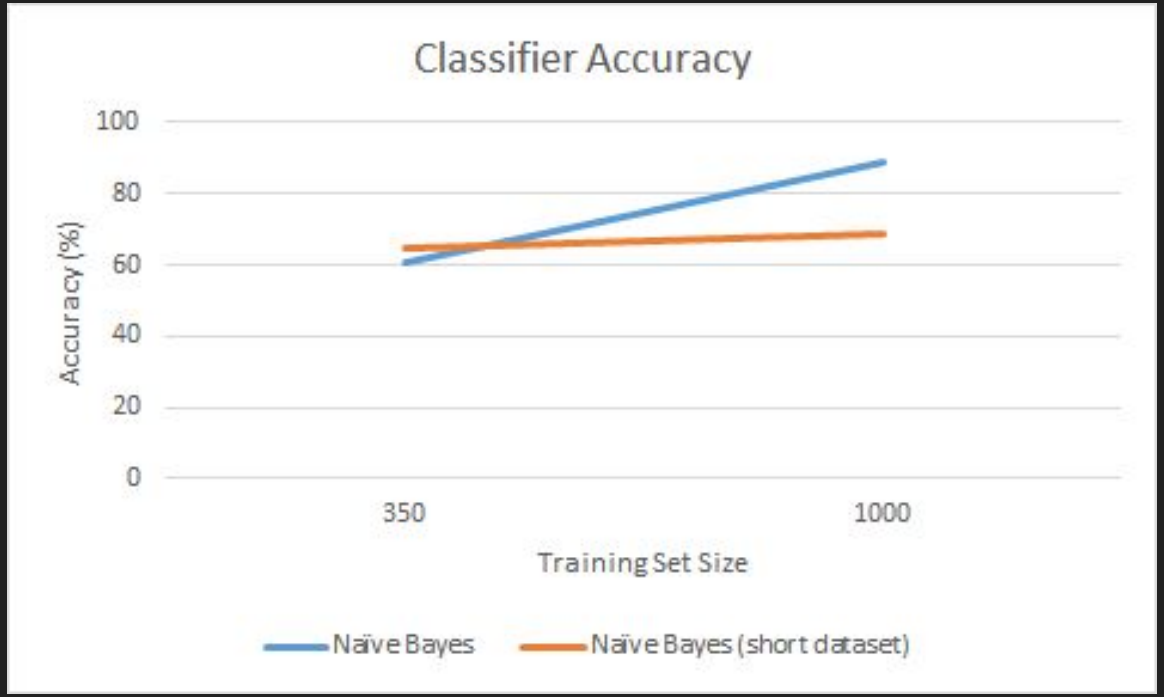
1. Naive Bayes classifier with full training set
2. Naive Bayes classifier with short training set
3. K-Nearest Neighbors classifier with short training set

Prototype and Testing - Results



Final Results

Training set increased to 1,000 records



Final Results

Classifier Used	Naive Bayes
Test Set Size	1,000
Amount of Complaints Classified as Recalls	1.5%
% of Complaints DB that are Recalls	1.5%
Classifier Accuracy	88%
Program Runtime	~2 minutes

QUESTIONS?