

Intro to MP 1

Lecture 3: **Intro to MP 1 and Jupyter Notebook**

ECE/CS 498 DS
Professor Ravi K. Iyer
University of Illinois

Announcements

- Office hours: Today from 4-5 PM in CSL 249
- Discussion section: The discussion this Friday 1/31 (4-5 PM, CSL 141) will serve as an additional office hour with the TAs
- **Submit your MP Groups by tonight @ 11:59 PM**
 - Signup via Google Form, link shared on Piazza (<https://forms.gle/nkjure9LXVjKKfEKA>)
 - Look at “Search for Partners” page on Piazza for help finding groups
 - Guidelines:
 - Groups should consist of exactly 3 students
 - You will work with the same groups for all the MPs (and if applicable, final project)
 - 4-credit hour students should only work with other 4 credit-hour students, and 3 credit-hour students should only work with other 3-credit hour students
- **HW0 has been released and it is due on 2/3 @ 11:59 PM**
 - Please upload your solved HW on Compass2G – **Let TAs know ASAP if you don't have access to Compass2G!**
 - You can either (i) handwrite and scan or (ii) type your answers

Mini-Project 1

Autonomous Vehicle Safety Analysis

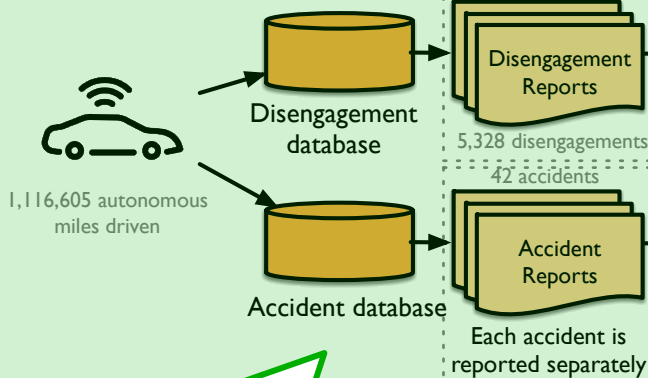
- You will learn and apply the following concepts:
 - Handling datasets
 - Basic statistical analysis of the dataset
 - Probabilistic analysis of the data using concepts from ECE313
 - Create a Naïve Bayes Machine Learning model to classify data
- Timeline
 - Release: **Thursday Jan 30, 2020.**
 - Checkpoint (Tasks 0-1): **Thursday February 6, 2020**
 - Final Submission (Tasks 0-3): **Thursday Feb 20, 2020**
 - Presentations: TBD (sometime soon after final submission...)
- Submissions
 - For each of checkpoint and final submission, please submit the following:
 - Single .ipynb file with all the relevant code
 - Powerpoint slides with answers to questions (we will provide template)
 - One submission per group
 - **Late submission policy will be applied**
- Individual contributions will be evaluated!

An End-to-End Workflow for AV Log Data Analysis

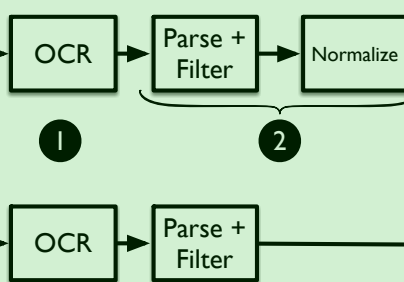
- Reports stored as scanned documents.
- Vendor specific parsing & filtering.
- Standardizing data formats across vendors.

- Analyze failure data to quantify
 - Causes
 - Dynamics
 - Impacts

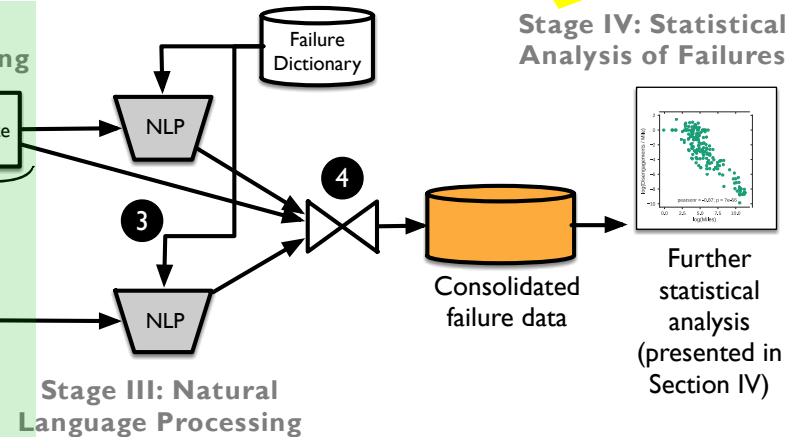
Stage I: Data Collection



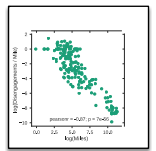
Stage II: Parsing and Filtering



Stage III: Natural Language Processing



Stage IV: Statistical Analysis of Failures



Further statistical analysis (presented in Section IV)

- Vendors are required to collect data as per CA laws.
- CA DMV curates databases of vendor reports.
- No standardized reporting formats.

- Localize failures in abstract system model.

Mini-Project 1

Dataset Description

You will be given two files.

- **mp1_av_disengagement.csv**

- This file lists the details of each disengagement that happened in AV testing.

| | Month | Car | Location | Weather | TypeOfTrigger | ReactionTime | Cause |
|---|--------|-------------------|--------------|---------|---------------|--------------|---------------------------------|
| 0 | 14-Sep | WDD2221851A001072 | urban-street | cloudy | automatic | 0.58 | Hardware Fault |
| 1 | 14-Sep | WDD2221851A001072 | highway | cloudy | manual | - | Unwanted Driver Discomfort |
| 2 | 14-Sep | WDD2221851A001072 | urban-street | clear | automatic | 1.93 | Adverse road surface conditions |

- Note: A **disengagement** occurs when a failure in the AV system causes control of the vehicle to switch from the software to the human driver

Mini-Project 1

Dataset Description

You will be given two files.

- **mp1_av_totalmiles.csv**
 - This file contains the total number of miles driven and other summary statistics by month.

| Month | Car | miles driven in autonomous mode | total number of disengagements | number of automatic disengagements | number of manual disengagements |
|--------|-------------------|--|-----------------------------------|--|------------------------------------|
| 14-Sep | WDD2221851A001072 | 4207 | 44 | 24 | 20 |
| 14-Oct | WDD2221851A001072 | 23971 | 228 | 103 | 125 |
| 14-Nov | WDD2221851A001072 | 15837 | 141 | 67 | 74 |

What questions would you ask from the data?

| | Month | Car | Location | Weather | TypeOfTrigger | ReactionTime | Cause |
|---|--------|-------------------|--------------|---------|---------------|--------------|---------------------------------|
| 0 | 14-Sep | WDD2221851A001072 | urban-street | cloudy | automatic | 0.58 | Hardware Fault |
| 1 | 14-Sep | WDD2221851A001072 | highway | cloudy | manual | - | Unwanted Driver Discomfort |
| 2 | 14-Sep | WDD2221851A001072 | urban-street | clear | automatic | 1.93 | Adverse road surface conditions |

Table 1

| Month | Car | miles driven in autonomous mode | total number of disengagements | number of automatic disengagements | number of manual disengagements |
|--------|-------------------|---------------------------------|--------------------------------|------------------------------------|---------------------------------|
| 14-Sep | WDD2221851A001072 | 4207 | 44 | 24 | 20 |
| 14-Oct | WDD2221851A001072 | 23971 | 228 | 103 | 125 |
| 14-Nov | WDD2221851A001072 | 15837 | 141 | 67 | 74 |

Table 2

Mini-Project 1

Task Descriptions

- **Task 0 (Getting to know the analysis environment)**
 - Before any analysis can be done on a given dataset, you will need to know how to import, handle and do some basic data manipulation programmatically. This task is designed to help you get accustomed to the Python Jupyter Notebook data analysis environment.
- **Task 1 (Basic analysis of AV disengagements)**
 - In this task, you will be performing some basic analysis on the datasets. It involves fitting probability distributions to the data and interpreting the distributions.

Mini-Project 1

Task Descriptions

- **Task 2 (Probabilistic analysis of AV disengagements)**
 - You will apply concepts like conditional probabilities, hypothesis testing and other probability theorems to extract meaningful insights of the safety of AVs from the provided datasets.
- **Task 3 (Using the Naïve Bayes Model)**
 - In this task, you will train a Naïve Bayes Machine Learning model based on the features (e.g. Location, Weather) in the dataset. After training, the model will be used to classify/predict the cause of an AV disengagement given a set of input features.