# Esame Machine Learning

Creo una cartella per il progetto

Faccio il venv 🡪 ok

python3 -m venv venv

creo il file .gitignore 🡪 ok

mi collego a github 🡪

git init

git add .

git commit -m "Primo commit"

git remote add origin https://github.com/MarcoBa001/project\_car\_acc.git

git push -u origin master  # o main, a seconda del nome del branch

git branch

git remote -v

ctrl + c per interrompere il caricamento di un push

Come scrivere il markdown

# Titolo di livello 1

## Titolo di livello 2

### Titolo di livello 3

#### Titolo di livello 4

##### Titolo di livello 5

###### Titolo di livello 6

3. Una breve presentazione in pdf (da 5 a 10 slide) che contenga:

# a. Titolo del progetto

con composizione del gruppo,

1 persona

fonte (URL) del dataset e

https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents

link al notebook

[US-Accidents: A Countrywide Traffic Accident Dataset - Sobhan Moosavi](https://smoosavi.org/datasets/us_accidents)

[[1906.05409] A Countrywide Traffic Accident Dataset](https://arxiv.org/abs/1906.05409)

[A Countrywide Traffic Accident Dataset](https://arxiv.org/pdf/1906.05409)

**US-Accidents: A Countrywide Traffic Accident Dataset**

**Published:** January 2021

**Description**

This is a countrywide traffic accident dataset, which covers 49 states of the United States. The data is continuously being collected from February 2016, using several data providers, including multiple APIs that provide streaming traffic event data. These APIs broadcast traffic events captured by a variety of entities, such as the US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks. Currently, there are about **1.5 million** accident records in this dataset. Check the below descriptions for more detailed information.

For any inquiries, contact me at moosavi.3@osu.edu

**Acknowledgments**

Please cite the following papers if you use this dataset:

* Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. [“A Countrywide Traffic Accident Dataset.”](https://arxiv.org/abs/1906.05409), arXiv preprint arXiv:1906.05409 (2019).
* Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. [“Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights.”](https://arxiv.org/abs/1909.09638) In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.

**Usage Policy and Legal Disclaimer**

This dataset is being distributed only for **Research** purposes, under [Creative Commons Attribution-Noncommercial-ShareAlike license (CC BY-NC-SA 4.0)](https://creativecommons.org/licenses/by-nc-sa/4.0/). By clicking on download button(s) below, you are agreeing to use this data only for non-commercial, research, or academic applications. You may cite the above papers if you use this dataset.

**Download**

* **Version 5 (Dec 2020):** In this version, we provide data that is collected from February 2016 to Dec 2020 for the [Contiguous United States](https://en.wikipedia.org/wiki/Contiguous_United_States). In this dataset, we have about **1.5 million** traffic accidents. Download data from [here](https://www.kaggle.com/sobhanmoosavi/us-accidents).

The next version will be available by December 2021.

**Coverage**

The dataset covers 49 states of the US. Following diagram shows the current data distribution over all the states.

**Format**

The data is provided in terms of a CSV file. Following table describes the data attributes (visit our [paper](https://arxiv.org/abs/1906.05409) to learn more about these attributes and how we obtained them):

| **#** | **Attribute** | **Description** | **Nullable** |
| --- | --- | --- | --- |
| 1 | ID | This is a unique identifier of the accident record. | No |
| 2 | Severity | Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay). | No |
| 3 | Start\_Time | Shows start time of the accident in local time zone. | No |
| 4 | End\_Time | Shows end time of the accident in local time zone. End time here refers to when the impact of accident on traffic flow was dismissed. | No |
| 5 | Start\_Lat | Shows latitude in GPS coordinate of the start point. | No |
| 6 | Start\_Lng | Shows longitude in GPS coordinate of the start point. | No |
| 7 | End\_Lat | Shows latitude in GPS coordinate of the end point. | Yes |
| 8 | End\_Lng | Shows longitude in GPS coordinate of the end point. | Yes |
| 9 | Distance(mi) | The length of the road extent affected by the accident. | No |
| 10 | Description | Shows natural language description of the accident. | No |
| 11 | Number | Shows the street number in address field. | Yes |
| 12 | Street | Shows the street name in address field. | Yes |
| 13 | Side | Shows the relative side of the street (Right/Left) in address field. | Yes |
| 14 | City | Shows the city in address field. | Yes |
| 15 | County | Shows the county in address field. | Yes |
| 16 | State | Shows the state in address field. | Yes |
| 17 | Zipcode | Shows the zipcode in address field. | Yes |
| 18 | Country | Shows the country in address field. | Yes |
| 19 | Timezone | Shows timezone based on the location of the accident (eastern, central, etc.). | Yes |
| 20 | Airport\_Code | Denotes an airport-based weather station which is the closest one to location of the accident. | Yes |
| 21 | Weather\_Timestamp | Shows the time-stamp of weather observation record (in local time). | Yes |
| 22 | Temperature(F) | Shows the temperature (in Fahrenheit). | Yes |
| 23 | Wind\_Chill(F) | Shows the wind chill (in Fahrenheit). | Yes |
| 24 | Humidity(%) | Shows the humidity (in percentage). | Yes |
| 25 | Pressure(in) | Shows the air pressure (in inches). | Yes |
| 26 | Visibility(mi) | Shows visibility (in miles). | Yes |
| 27 | Wind\_Direction | Shows wind direction. | Yes |
| 28 | Wind\_Speed(mph) | Shows wind speed (in miles per hour). | Yes |
| 29 | Precipitation(in) | Shows precipitation amount in inches, if there is any. | Yes |
| 30 | Weather\_Condition | Shows the weather condition (rain, snow, thunderstorm, fog, etc.) | Yes |
| 31 | Amenity | A [POI](https://wiki.openstreetmap.org/wiki/Points_of_interest) annotation which indicates presence of [amenity](https://wiki.openstreetmap.org/wiki/Key:amenity) in a nearby location. | No |
| 32 | Bump | A POI annotation which indicates presence of speed bump or hump in a nearby location. | No |
| 33 | Crossing | A POI annotation which indicates presence of [crossing](https://wiki.openstreetmap.org/wiki/Key:crossing) in a nearby location. | No |
| 34 | Give\_Way | A POI annotation which indicates presence of [give\_way](https://wiki.openstreetmap.org/wiki/Tag:highway%3Dgive_way) in a nearby location. | No |
| 35 | Junction | A POI annotation which indicates presence of [junction](https://wiki.openstreetmap.org/wiki/Key:junction) in a nearby location. | No |
| 36 | No\_Exit | A POI annotation which indicates presence of [no\_exit](https://wiki.openstreetmap.org/wiki/Key:noexit) in a nearby location. | No |
| 37 | Railway | A POI annotation which indicates presence of [railway](https://wiki.openstreetmap.org/wiki/Key:railway) in a nearby location. | No |
| 38 | Roundabout | A POI annotation which indicates presence of [roundabout](https://wiki.openstreetmap.org/wiki/Tag:junction%3Droundabout) in a nearby location. | No |
| 39 | Station | A POI annotation which indicates presence of [station](https://wiki.openstreetmap.org/wiki/Key:station) in a nearby location. | No |
| 40 | Stop | A POI annotation which indicates presence of [stop](https://wiki.openstreetmap.org/wiki/Key:stop) in a nearby location. | No |
| 41 | Traffic\_Calming | A POI annotation which indicates presence of [traffic\_calming](https://wiki.openstreetmap.org/wiki/Key:traffic_calming) in a nearby location. | No |
| 42 | Traffic\_Signal | A POI annotation which indicates presence of [traffic\_signal](https://wiki.openstreetmap.org/wiki/Tag:highway%3Dtraffic_signals) in a nearby loction. | No |
| 43 | Turning\_Loop | A POI annotation which indicates presence of [turning\_loop](https://wiki.openstreetmap.org/wiki/Tag:highway%3Dturning_loop) in a nearby location. | No |
| 44 | Sunrise\_Sunset | Shows the period of day (i.e. day or night) based on sunrise/sunset. | Yes |
| 45 | Civil\_Twilight | Shows the period of day (i.e. day or night) based on [civil twilight](https://en.wikipedia.org/wiki/Twilight#Civil_twilight). | Yes |
| 46 | Nautical\_Twilight | Shows the period of day (i.e. day or night) based on [nautical twilight](https://en.wikipedia.org/wiki/Twilight#Nautical_twilight). | Yes |
| 47 | Astronomical\_Twilight | Shows the period of day (i.e. day or night) based on [astronomical twilight](https://en.wikipedia.org/wiki/Twilight#Astronomical_twilight). | Yes |

# b. Data exploration

# c. Data preprocessing

# d. Clustering del dataset (escludendo l’etichetta di classe)

i. Algoritmo usato e giustificazione della scelta

ii. Descrizione dei cluster ottenuti anche usando grafici

e. Classificazione

i. Algoritmi di classificazione usati (almeno 2)

ii. Performance evaluation (matrici di confusione, accuracy, f1 score,

…)

iii. Commento sui risultati

f. Idee per sviluppi futuri (come migliorare i risultati)

# DATA MINING PROCESS

1. Learning the application domain:

– relevant prior knowledge and goals of application

2. Creating a target data set: data selection

3. Data cleaning and preprocessing (may take 60% of effort!)

4. Data reduction and transformation:

– Find useful features, dimensionality/variable reduction, invariant

representation.

5. Choosing functions of data mining (summarization, classification,

regression, association, clustering)

6. Choosing the mining algorithm(s)

7. Data mining: search for patterns of interest

8. Pattern evaluation and knowledge presentation

– visualization, transformation, removing redundant patterns, etc.

9. Use of discovered knowledge