

Data

To solve the objective, we need following data: -

1. List of all the accidents occurred and recorded in the Seattle.
2. Weather condition during the travel for every recorded accident.
3. Lighting condition on the Road for every recorded accident.
4. Road condition for every recorded accident.
5. Before the accident, whether the victim was under the influence of alcohol/substance.
6. Whether it was violation of speed limit.

Sources of Data

The data has been made available by the IBM online certification team. The collisions data is provided by Seattle Police Department and recorded by Traffic Records which is being updated on weekly basis since 2004.

Approach

There are 6 steps to build a model –

1. Business understanding

The initial phase is to understand the project's objective from the business or application perspective. Then, you need to translate this knowledge into a machine learning problem with a preliminary plan to achieve the objectives.

2. Data understanding

In this phase, you need to collect or extract the dataset from various sources such as csv file or SQL database. Then, you need to determine the attributes (columns) that you will use to train your machine learning model. Also, you will assess the condition of chosen attributes by looking for trends, certain patterns, skewed information, correlations, and so on.

3. Data preparation and cleaning

The data preparation includes all the required activities to construct the final dataset which will be fed into the modeling tools. Data preparation can be performed multiple times and it includes balancing the labeled data, transformation, filling missing data, and cleaning the dataset.

4. Modelling

In this phase, various algorithms and methods can be selected and applied to build the model including supervised machine learning techniques. You can select k Nearest Neighbor, SVM, XGBoost, decision tree, or any other techniques. You can select a single or multiple machine learning models for the same data mining problem. At this phase, stepping back to the data preparation phase is often required.

5. Evaluation

Before proceeding to the deployment stage, the model needs to be evaluated thoroughly to ensure that the business or the applications' objectives are achieved. Certain metrics can be used for the model evaluation such as accuracy, recall, F1-score, precision, and others.

6. Deployment

The deployment phase requirements vary from project to project. It can be as simple as creating a report, developing interactive visualization, or making the machine learning model available in the production environment. In this environment, the customers or end-users can utilize the model in different ways such as API, website, or so on.