**Project Documentation: Accident Dataset Decision Tree Analysis using WEKA**

**1. Objective:**

The goal of this analysis is to predict accident severity using a decision tree classifier in **WEKA**. The dataset contains information related to road accidents, including vehicle type, driver demographics, weather conditions, and road conditions. The decision tree model helps identify the most important factors that influence the severity of accidents.

**2. Methodology:**

**Step 1: Dataset Loading**

* The dataset, located at C:\Users\admin\Desktop\WORK\datasets\Kaagle\_Upload.csv, was first prepared and loaded into WEKA.
* The CSV file was imported into WEKA’s **Explorer** interface using the "Open file" option under the **Preprocess** tab.

**Step 2: Data Preprocessing**

* After loading the dataset into WEKA, irrelevant columns were excluded, and missing values were handled using WEKA’s built-in preprocessing filters.
  + **Relevant Columns**:
    - vehicle\_type: The type of vehicle involved in the accident.
    - vehicle\_manoeuvre: The vehicle's action during the accident.
    - sex\_of\_driver: The gender of the driver.
    - age\_of\_driver: The age of the driver.
    - road\_conditions: Condition of the road during the accident.
    - weather\_conditions: The weather at the time of the accident.
    - casualty\_severity: The severity of the accident (target variable).
* We used WEKA’s **Replace Missing Values** filter under the **Filter** section to handle missing data.

**Step 3: Splitting the Data**

* The dataset was split into **training** and **testing** sets:
  + **70% Training Set**: Used to train the decision tree model.
  + **30% Testing Set**: Used for evaluating the model’s performance.
* The split was done in the **Classify** tab, using the "Percentage Split" option to separate the data into training and testing sets.

**Step 4: Decision Tree Algorithm Selection**

* WEKA’s **J48** algorithm (an implementation of the C4.5 decision tree) was selected for classification.
  + The J48 classifier was chosen due to its simplicity and interpretability for decision-making processes.
* The default settings of J48 were used, including the **confidence factor** of 0.25 and **minimum number of instances per leaf** of 2.

**Step 5: Training the Model**

* In the **Classify** tab, the J48 algorithm was applied to the training set, generating a decision tree model.
* The tree was constructed based on the various attributes, such as vehicle type, weather conditions, and road conditions, to predict the severity of an accident.

**Step 6: Model Evaluation**

* The model was evaluated using the **test set** (30% of the data) by applying the "Test on test split" option.
* WEKA provided several evaluation metrics:
  + **Accuracy**: The percentage of correctly classified instances.
  + **Confusion Matrix**: Showed the distribution of predicted versus actual classes.
  + **Precision, Recall, F-Measure**: Used to assess the quality of predictions for each class of accident severity.

**Step 7: Visualizing the Decision Tree**

* The decision tree generated by the J48 algorithm was visualized in WEKA, providing an easy-to-understand graphical representation of how different features influence accident severity.
* The decision tree could be viewed by selecting "Visualize tree" after the classification process.

**3. Insights from the Decision Tree Analysis:**

1. **Key Factors Influencing Accident Severity**:
   * The decision tree highlighted that **road conditions** and **weather conditions** were critical factors influencing accident severity. Poor road conditions combined with adverse weather led to more severe accidents.
   * **Vehicle type** also played a role, with larger vehicles often involved in more serious accidents.
2. **Road and Weather Conditions**:
   * Accidents were more likely to be severe when the weather was unfavorable (e.g., rain, fog, or icy conditions) and the roads were in poor condition.
   * Insights suggest that road safety improvements during bad weather conditions can reduce the occurrence of severe accidents.
3. **Driver Demographics**:
   * The decision tree also indicated that younger or older drivers, particularly in specific vehicle types, were more likely to be involved in severe accidents.
   * This suggests a potential need for targeted road safety campaigns based on age demographics.
4. **Actionable Insights**:
   * Focus on improving road safety during adverse weather and educating drivers on proper driving techniques during poor weather conditions.
   * Policymakers could also use this information to implement stricter regulations for certain vehicle types or road conditions.

**4. Conclusion:**

Using WEKA’s J48 decision tree algorithm, we were able to identify important factors that contribute to accident severity. The most influential factors included road conditions, weather conditions, and vehicle type. By analyzing these attributes, we gained valuable insights into how accidents occur and the factors that increase the severity. These findings can inform road safety measures, such as improving road infrastructure in hazardous conditions and promoting safer driving behaviors in bad weather.