Titanic Survival Prediction using NAIVE BAYES

1. Finding the Problem - Application
   - To Predict: Titanic Survival prediction based on some parametes
2. Collecting Dataset
   - Input: Age,Pclass, Sex & Fare Output: Survived or Not
   - Based on Age,Pclass, Sex & Fare, we gonna predict whether the person might survived or not
3. Load & Summarize Dataset
   - Load Dataset from the directory & Summarize the details such as no. of rows and Columns & Content
   - Pandas - Load CSV Format Dataset: dataset = pandas.read_csv('dataset.csv')
   - No. of Rows & Columns: dataset.shape
   - Display 1st 5 rows of dataset: dataset.head(5)
4. Mapping Gender text data to Binary Numbers
   - Function: .map
   - If the data is male then it is gonna be mapped to 1, female to 0
5. Segregating Dataset into X & Y
   - dataset
   - dataset.drop - To remove column: X = dataset.drop('Survived',axis='columns')
   - Pic the certain column: Y = dataset.Survived
6. Splitting Dataset to Train & Test
   - Useful for validation
   - train_test_split(X, Y, test_size = 0.25, random_state = 0)
7. Identify & Replacing NaN values in Dataset
   - Checking for Nan: X.isna().any()
   - Replacing NaN values with Mean of that certain feature: X.Age.fillna(X.Age.mean())
8. Algorithm
   - naive_bayes - Based on Bayes Theorem
   - $P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$
   - Every pair of features being classified is independent of each other & having Equal contribution
   - BAYES THEOREM
     - probability of an event occurring given the probability of another event that has already occurred
       1. To Find: I pick Diamond Card, Need to Find probability of that card is Queen
       2. Probability of finding Queen in total cards: P(Queen)=4/52=1/13
       3. Probability of finding Diamonds from Total cards: P(Diamonds)=13/52=1/4
       4. Probability of finding Diamond in the Queen: P(Diamond/Queen) = 1/4
       5. $P(Queen/Diamond) = \frac{P(Diamond/Queen) * P(queen)}{P(Diamond)}$ — Solution: (1/4*1/13)/(1/4) = 1/13
   - Types
     - Gaussian Naive Bayes
       - $P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$
       - The parameters and are estimated using maximum likelihood
       - The likelihood of the features is assumed to be Gaussian
     - Multinomial Naive Bayes
     - Complement Naive Bayes
     - Bernoulli Naive Bayes
     - Categorical Naive Bayes
9. Training
   - Training our Model for Pre-processed Dataset
   - model.fit(X_train, y_train)
10. Validation
    - Obtaining the accuracy of the Model
    - Accuracy Score
11. Prediction
    - Observing how our model is classifying our new data
    - result = model.predict(personData)