AL2002 – Artificial Intelligence Lab

Lab Task # 07

Note:

- Plagiarism will not be **tolerated!!**
- Use comments wherever applicable.
- Please ensure to submit both a PDF document and a Python file containing your code on the classroom platform.

Problem: 1 - Predicting Titanic Survival with KNN

Download the **Titanic** dataset from the internet, which contains information about passengers onboard the Titanic. You're interested in predicting survival outcomes using a **k-Nearest Neighbors (KNN)** machine learning model. The features you plan to use are "Pclass" (Passenger Class), "Gender," "Age," "SibSp" (Number of Siblings/Spouses Aboard), "Parch" (Number of Parents/Children Aboard), "Fare," and "Embarked" (Port of Embarkation). The target variable you want to predict is "Survived."

Since KNN works with numerical data, you need to convert categorical features like "Gender" and "Embarked" into numeric format using feature engineering techniques.

1. Describe how you would convert categorical features like "Gender" and "Embarked" into numeric format.
2. Implement the KNN model using the features mentioned above and the target variable "Survived." Apply feature engineering techniques as described in part 1.
3. Perform the following steps:
    - Split the dataset into training and testing sets.
    - Train the KNN model on the training set.
    - Evaluate the model's accuracy on the testing set.
    - Repeat steps 1-3 for different random states (e.g., 1, 10, 42) to observe how the accuracy varies.
4. Plot a graph showing the accuracy of the KNN model on the testing set for different random states.

Make sure to interpret the results and discuss any observations you make regarding the accuracy variations.