# Load CSV Data

**Project Overview: Titanic Data Classification**

This project involves building a deep learning model to predict whether passengers survived the Titanic disaster using a neural network.

**Steps:**

1. **Data Loading**:
   * Used pandas to load Titanic dataset.
   * Split data into features (age, fare, etc.) and labels (survived or not).
   * Preprocessed data by converting categorical variables to numerical and normalizing continuous features.
2. **Model Building**:
   * Used TensorFlow and Keras for building the model.
   * Model architecture includes:
     + Input layer for features.
     + Dense layers with ReLU activation.
     + Output layer with 1 neuron (for binary classification).
3. **Model Compilation**:
   * Compiled the model using BinaryCrossentropy loss and Adam optimizer.
4. **Model Training**:
   * Trained the model on the Titanic dataset for 10 epochs, using the fit() method.
   * Monitored loss and accuracy during training.
5. **Model Evaluation**:
   * Evaluated model performance on test data.
   * Measured accuracy and loss on the test dataset.
6. **Results**:
   * Achieved a test accuracy of ~33.33% (indicating potential for improvement).

**Key Concepts:**

* **Preprocessing**: Converting categorical data (e.g., sex, class) into numeric and normalizing continuous values.
* **Neural Network Layers**: Used a simple feedforward neural network with dense layers.
* **TensorFlow/Keras**: Used TensorFlow/Keras for model creation, training, and evaluation.
* **Loss Function**: BinaryCrossentropy for binary classification tasks.
* **Optimizer**: Adam optimizer for efficient training.

**project like the Titanic Survival Prediction Model:**

**1. What is the Titanic dataset?**

* **Answer**: The Titanic dataset contains information about passengers aboard the Titanic. It includes features such as age, class, fare, and whether the passenger survived. The goal is to predict if a passenger survived based on these features.

**2. How do you handle missing data in a dataset?**

* **Answer**: You can handle missing data by:
  + Dropping rows or columns with missing values.
  + Filling missing values using techniques like mean, median, or mode imputation.
  + Using machine learning models to predict missing values.

**3. What are activation functions in neural networks?**

* **Answer**: Activation functions introduce non-linearity into the model, allowing it to learn complex patterns. Common activation functions include:
  + **ReLU** (Rectified Linear Unit) for hidden layers.
  + **Sigmoid** or **Softmax** for the output layer in classification tasks.

**4. Explain the difference between training and testing data.**

* **Answer**: Training data is used to train the model, while testing data is used to evaluate the model's performance on unseen data. It's important to split the data to avoid overfitting.

**5. What is the purpose of a loss function in neural networks?**

* **Answer**: The loss function measures how well the model’s predictions match the actual results. In binary classification, **Binary Crossentropy** is commonly used.

**6. What is the Adam optimizer, and why is it commonly used?**

* **Answer**: Adam (Adaptive Moment Estimation) is an optimizer that combines the benefits of two other extensions of stochastic gradient descent. It's efficient and handles sparse gradients well, making it ideal for many machine learning tasks.

**7. What is overfitting, and how can you prevent it?**

* **Answer**: Overfitting occurs when a model learns the training data too well, including noise, which reduces its performance on new data. It can be prevented by:
  + Using more data.
  + Implementing early stopping.
  + Using regularization (L1/L2).
  + Cross-validation.

**8. What is the purpose of using the fit() function in Keras?**

* **Answer**: The fit() function is used to train the model on the data. It takes the training data, labels, number of epochs, batch size, and other parameters like validation data.

**9. What does the evaluation function evaluate() do in Keras?**

* **Answer**: The evaluate() function computes the loss and metrics (like accuracy) on the test dataset after the model has been trained.

**10. How do you handle categorical data?**

* **Answer**: Categorical data can be handled by:
  + One-Hot Encoding: Creating binary columns for each category.
  + Label Encoding: Assigning a unique integer to each category.

**11. What is the purpose of normalization and scaling in machine learning?**

* **Answer**: Normalization or scaling ensures that all features contribute equally to the model. This is especially important for models like neural networks that use gradient descent.

**12. What is the role of a model’s architecture?**

* **Answer**: The architecture of a model defines its structure, including the number of layers, number of neurons in each layer, activation functions, etc. The right architecture depends on the problem you are trying to solve.