Titanic Survival Prediction

Predicting Survival on the Titanic using Passenger Information

Ram Sunder

Contents

Introduction

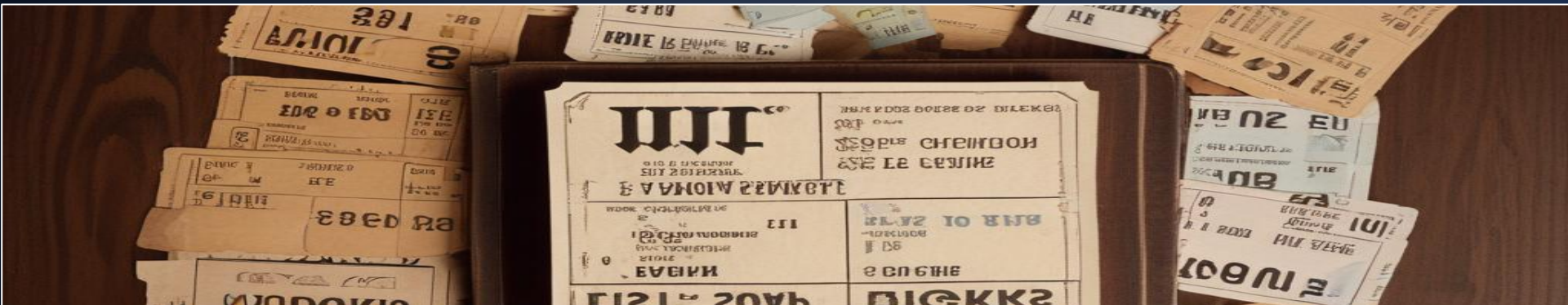
Use Case:

- The Titanic dataset contains information about passengers on the Titanic.
- Goal: Predict whether a passenger survived the disaster based on features like age, gender, class, etc.
- Significance: Helps in understanding factors that could affect survival rates and assists in building predictive models.
- Approach: We build an Artificial Neural Network (ANN) to predict whether a passenger survived based on their demographic and travel-related features.

Dataset Overview

Dataset: Titanic dataset, containing 891 rows and 69 features including passenger details such as Age, Sex, Fare, Embarked, etc.
Target Variable: 'Survived' (0 = Did not survive, 1 = Survived)
Features:

- Categorical Features: 'Sex' and 'Embarked' are encoded as 0 and 1.
- Numerical Features: 'Age', 'Fare' are already scaled and cleaned.
- No Missing Values: All missing data points have been addressed (already pre-processed).



Features of the Titanic Dataset

Key Features:

- Survived: Whether the passenger survived (1) or not (0).
- Pclass: The passenger's class (1st, 2nd, 3rd).
- Sex: Gender of the passenger.
- Age: Age of the passenger.
- SibSp: The number of siblings or spouses aboard.
- Parch: The number of parents or children aboard.
- Fare: The ticket fare the passenger paid.
- Embarked: The port of embarkation (C = Cherbourg; Q = Queenstown; S = Southampton).

Technology used in EDA

Training and Evaluation

- The dataset was split into training and testing subsets, utilizing 80% of the data for training purposes and 20% for evaluation.

Algorithms Used

- Artificial Neural Network (ANN):
  - We used an ANN model with layers and activation functions to learn patterns from the Titanic dataset.
  - Activation Functions:
    - ReLU (Rectified Linear Unit) used in layers for non-linear transformations.
    - Sigmoid or Softmax used for output layer to predict binary survival outcome.

Key Concepts Explained

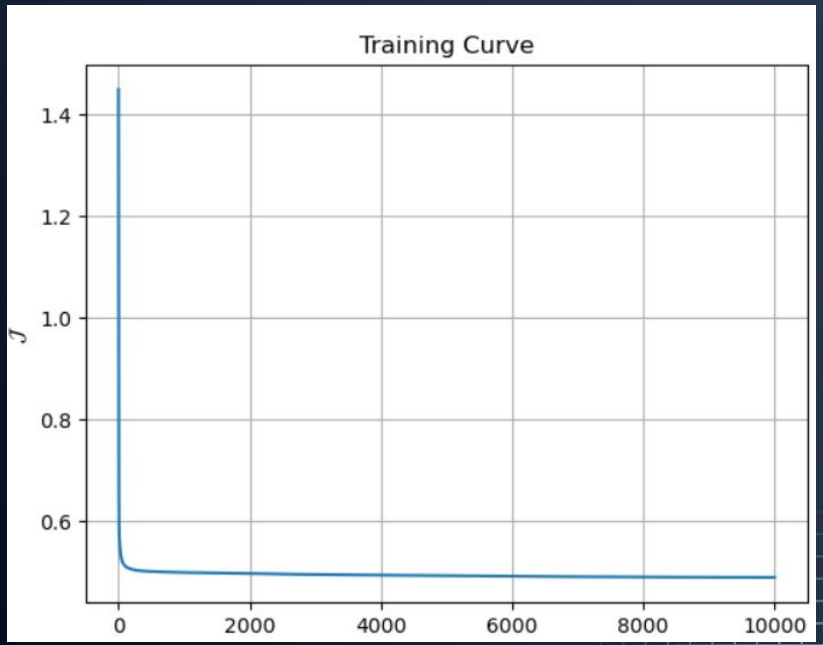
- Why Scaling?:
    - Features like Age and Fare have different scales. Scaling ensures that each feature contributes equally to the model, preventing certain features from dominating due to their larger values.
- Feature Encoding: Sex and Embarked are categorical features, encoded into numerical values (0, 1) for compatibility with the neural network.

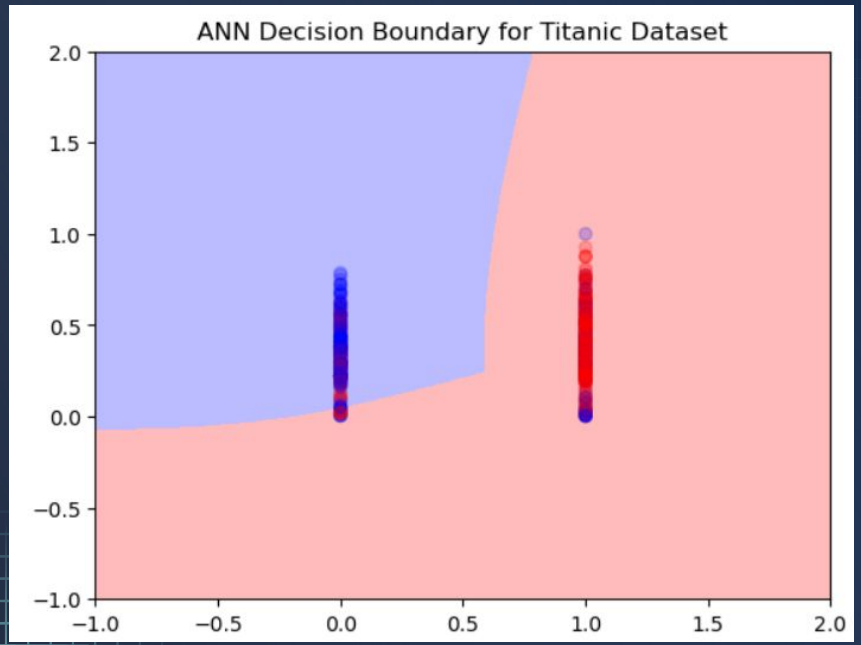
Output

- Model Prediction: The ANN predicts whether a passenger survived or not based on the features provided.
- Training Accuracy: Measures how well the model performs on the training dataset.

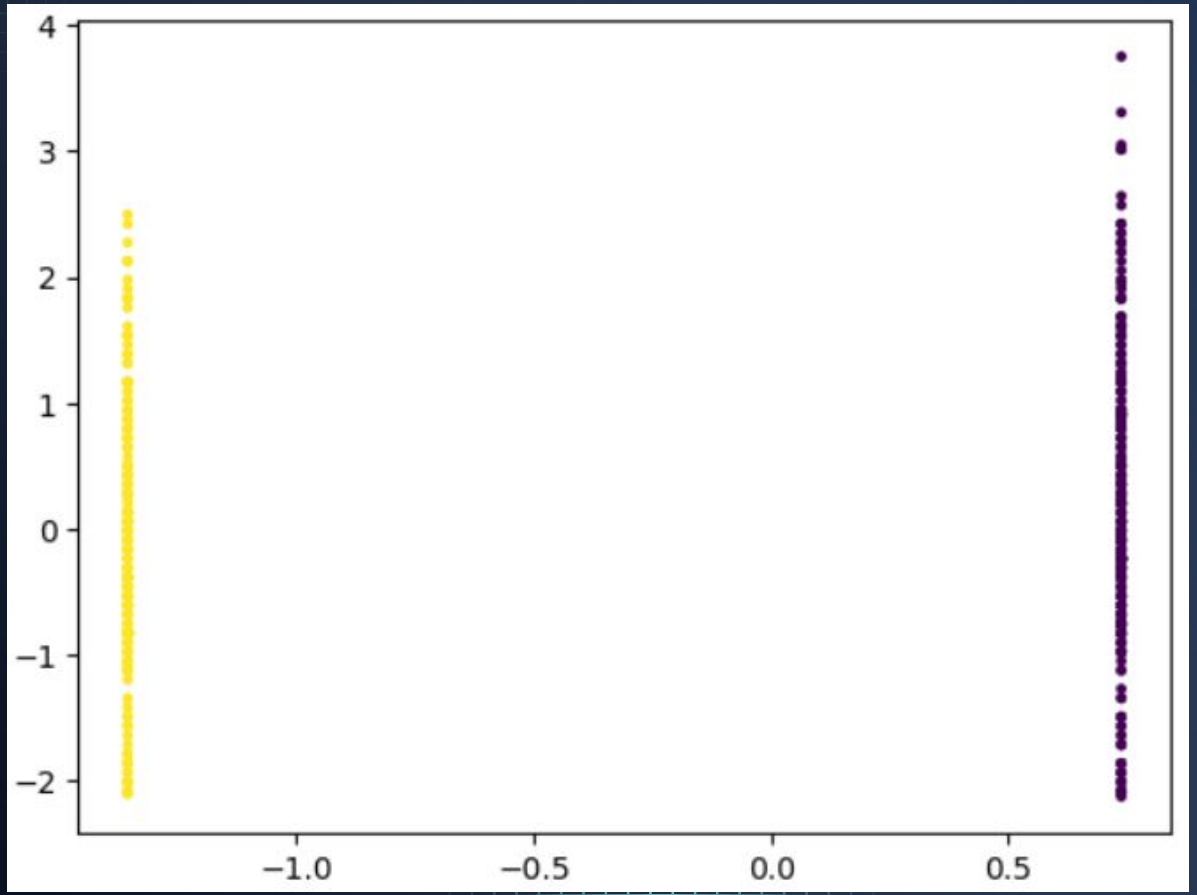
Visualizing the Decision Boundary

Training Accuracy: 0.7921





Scatterplot for Prediction



Conclusions

The predictive analysis of Titanic survival demonstrates the essential role of machine learning in deriving insights from historical datasets. Through careful model selection and evaluation, critical factors influencing survival were identified, enhancing our understanding of the event and informing future studies of similar nature.

Thank you!

Do you have any questions?