AI-LAB PROJECT

ZAIN ALI

BSAI-3A-046

PROJECT DOCUMENTATION

Titanic Survival Prediction Using Machine Learning

1. Introduction

This project predicts whether a passenger on the Titanic survived or not. We use machine learning to study the data and find out which factors affected survival, such as age, gender, ticket class, and fare.

2. Objective

- Build a survival prediction model.
- Clean and prepare the Titanic dataset.
- Identify important features.
- Evaluate model accuracy.

3. Dataset Description

Dataset contains:

- Survived (0/1)
- Pclass
- Name
- Sex
- Age
- SibSp
- Parch
- Ticket
- Fare
- Cabin

- Embarked

4. Tools & Technologies

- Python, Pandas, NumPy

- Matplotlib / Seaborn

- Scikit-learn

5. Data Preprocessing

- Filled missing Age with median.

- Filled missing Embarked with most common value.

- Removed Cabin (too many missing values).

- Dropped Name, Ticket, Cabin.

- Categorical encoding: Sex, Embarked.

- Train-test split: 80/20.

6. Machine Learning Model

Used Logistic Regression for binary classification.

7. Model Training

Model learns patterns such as:

- Females have higher survival rate.

- 1st class passengers survive more.

- Younger passengers survive more.

8. Model Evaluation

Accuracy: around 75%-85%

Confusion matrix: shows correct/incorrect predictions.

9. Results

- Females survived more.

- 1st class survival highest.

- Children survived more.

- Model gives good prediction accuracy.

10. Conclusion

Machine learning can predict Titanic survival with good accuracy. Major survival factors: gender, class, age.

11. Future Improvements

- Use Random Forest / XGBoost.
- Hyperparameter tuning.
- More visualizations.
- Try deep learning.

12. References

- Kaggle Titanic Dataset
- Python & Sklearn Documentation