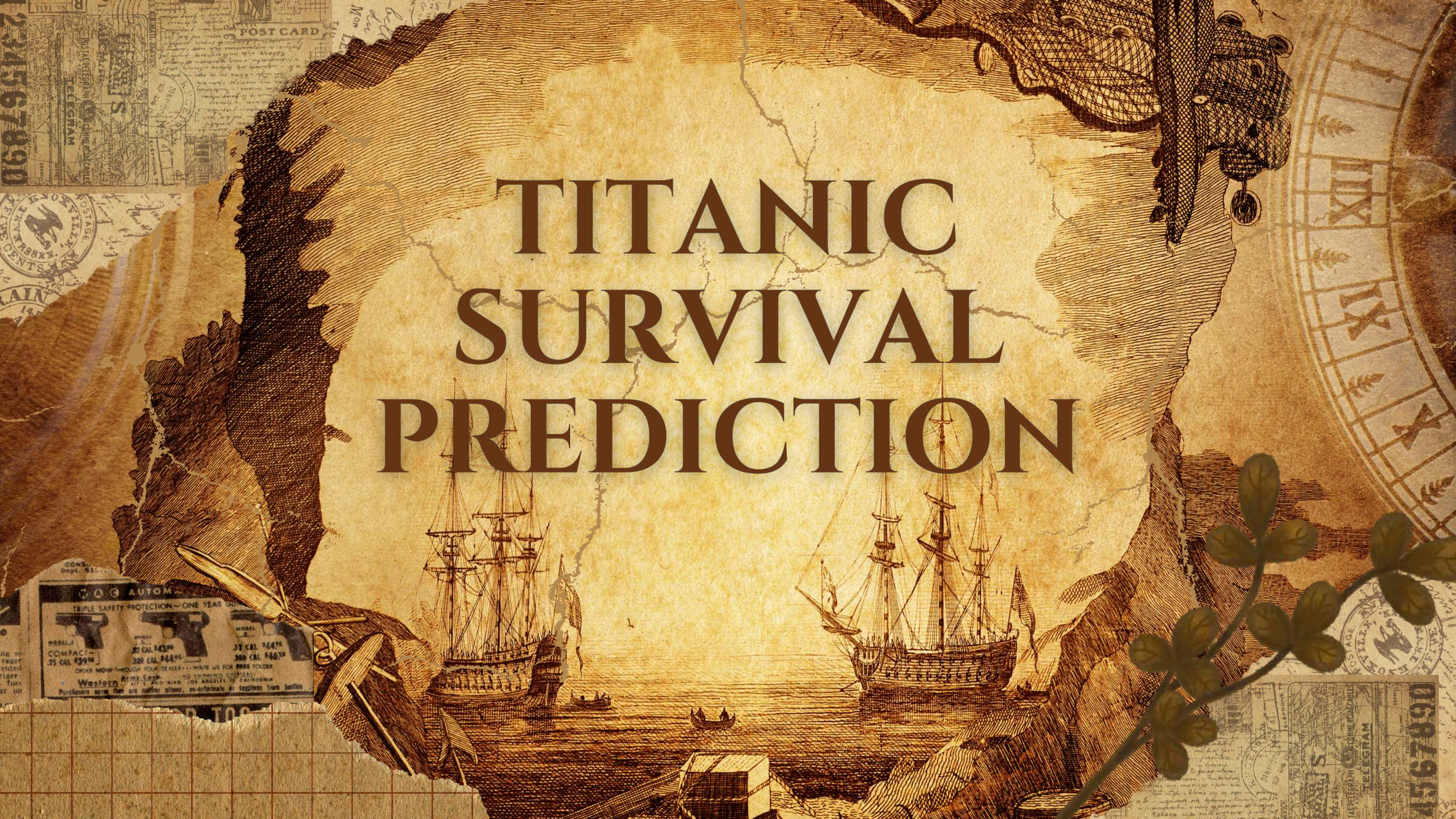
TITANIC
SURVIVAL
PREDICTION

TEAM MEMBERS

- Radhey Pal
- Rohan Sharma
- Sneh Singh
- Sudhanshu Singh
- Sushant Sharma

PROBLEM STATEMENT

- Objective: Predict whether a passenger survived the Titanic shipwreck based on a set of features using a machine learning classification model.
- Required: Data preprocessing and implementation of Naive Bayes classifier.
- Dataset Source: Kaggle Titanic Dataset https://www.kaggle.com/c/titanic

INTRODUCTION

- The sinking of the Titanic is one of the most infamous shipwrecks in history.
- Over 1500 passengers lost their lives.
- Kaggle's Titanic dataset challenges us to build a predictive model to determine survival probability based on passenger features.

DATASET OVERVIEW

- Train.csv: Contains 891 records with features and labels.
- Test.csv: Contains 418 records without survival labels.
- gender_submission.csv: Sample submission file.

IMPORTANT COLUMNS:

- PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked

DATA CLEANING

1. Missing Data Handling:
- Age → filled with median
- Embarked → filled with mode
- Cabin → dropped
2. Encoding:
- Sex: Male=0, Female=1
- Embarked: C=0, Q=1, S=2
3. Removed irrelevant: Name, Ticket, Cabin

FEATURE SELECTION

- <u>Used:</u> Pclass, Sex, Age, SibSp, Parch, Fare, Embarked
- <u>Dropped:</u> Name, Ticket, Cabin
- These selected features impact survival probability the most.

Why Naive Bayes?

- Easy to implement, fast, and good for small datasets.
- Works well with categorical variables.
- Assumes feature independence, which simplifies modeling.

MODEL BUILDING

- Used GaussianNB from sklearn:

from sklearn.naive_bayes import GaussianNB

```
from sklearn.naive_bayes import GaussianNB
model = GaussianNB()
model.fit(X_train, y_train)
predictions = model.predict(X_test)
```

- Trained on cleaned train.csv, predicted on test.csv.

Evaluation Metrics

- Accuracy: ~78% on training data
- Confusion Matrix: Shows TP, TN, FP, FN
- Classification Report:

  Precision, Recall, F1-Score

OUTPUT / RESULTS

- Submission file: PassengerId, Survived
- Observations:
- Females had higher survival rates
- 1st class passengers had better chances
- Children also had higher survival

CHALLENGES

- Missing values and sparse columns (Cabin)
- Small size of dataset
- Naive Bayes assumes independence, which may not hold true

CONCLUSION

- Built a clean, interpretable model using Naive Bayes
- Achieved good accuracy with basic preprocessing
- Found clear survival patterns in the data

FUTURE SCOPE

- Try ensemble models like Random Forest, XGBoost
- Tune hyperparameters with GridSearchCV
- Use feature interactions and visualizations

REFERENCES

- Titanic Dataset: https://www.kaggle.com/c/titanic
- Scikit-learn Docs: https://scikit-learn.org
- Python Libraries: pandas, numpy, seaborn, matplotlib


THANK
YOU