SVKM'S NMIMS Deemed to be UNIVERSITY | NAVI MUMBAI

By Anish and Moksha

LOGISTIC REGRESSION

Machine Learning

HELLO!

Warm greetings to all present. As we gather here today, I am excited to introduce our presentation in which we will predict the survival rate of people in titanic using logisitic regression.

AGENDA OVERVIEW

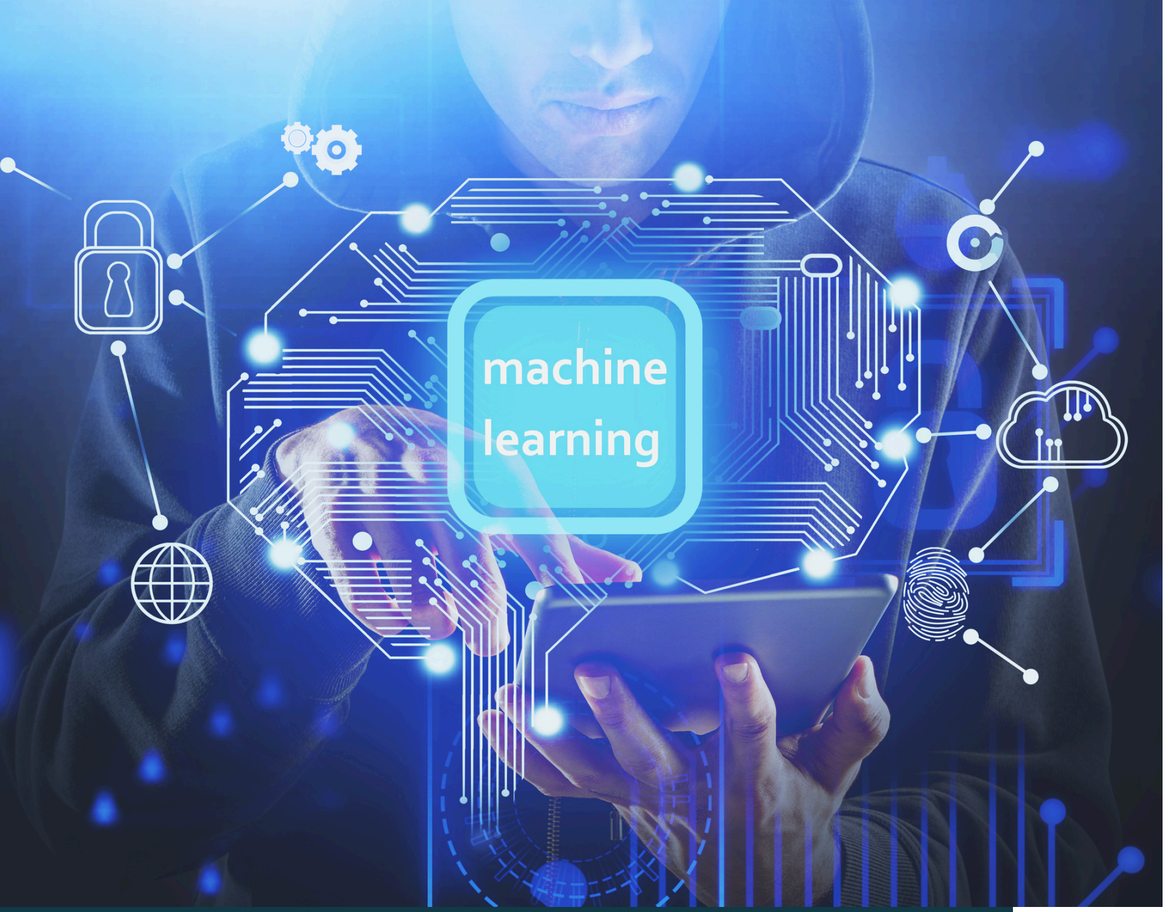
01 Introduction

02 Applications

03 Use cases

04 Model Fitting

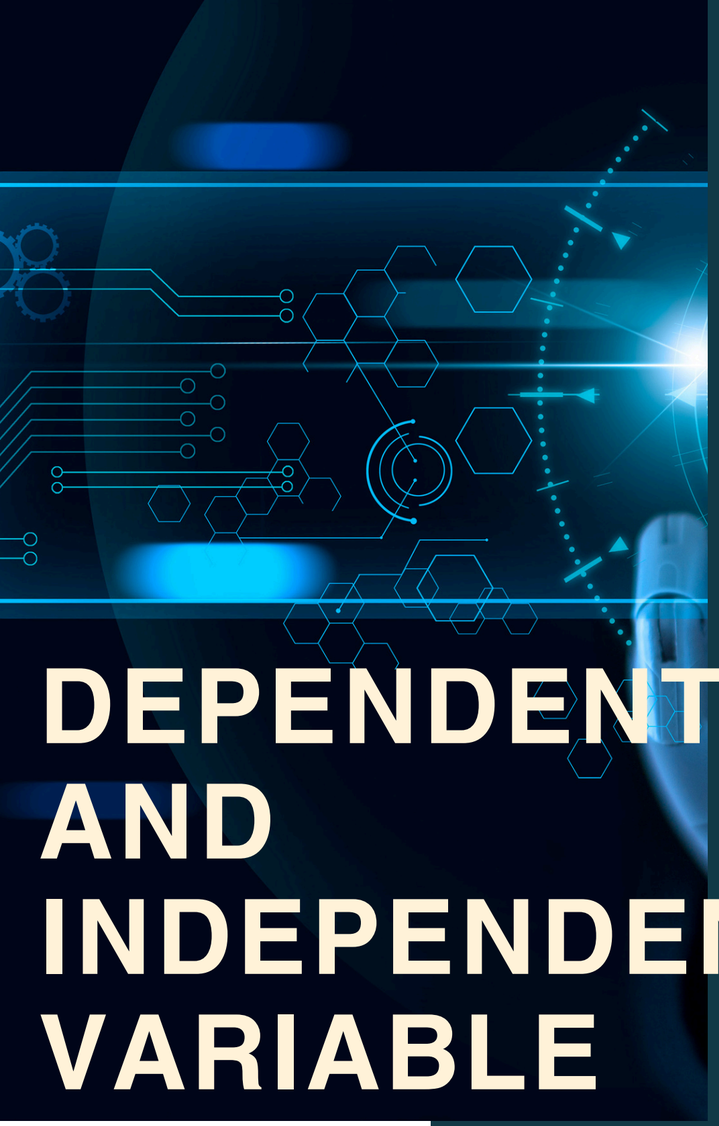
05 Conclusion



INTRODUCTION

01

Logistic regression is a powerful statistical technique used to model the relationship between a binary dependent variable and one or more independent variables. It is widely used in various fields, from healthcare to marketing, to make predictions and understand the factors influencing a specific outcome.



DEPENDENT AND INDEPENDENT VARIABLE

Dependent Variable

The dependent variable in logistic regression is a binary or dichotomous outcome, such as whether a person has a certain disease or not, or whether a customer will make a purchase or not.

Independent Variable

The independent variables, also known as predictors, are the factors that may influence the dependent variable. These can be continuous, categorical, or a combination of both.

02 APPLICATIONS

01

Disease Diagnosis

Logistic regression is used to predict the likelihood of a patient having a specific disease based on their symptoms and risk factors.

02

Credit Risk Assessment

Logistic regression models are employed to evaluate the creditworthiness of loan applicants and predict the probability of default.

03

Customer Churn Prediction

Logistic regression is applied to identify the factors that influence customer attrition and predict the likelihood of a customer leaving a service or product.

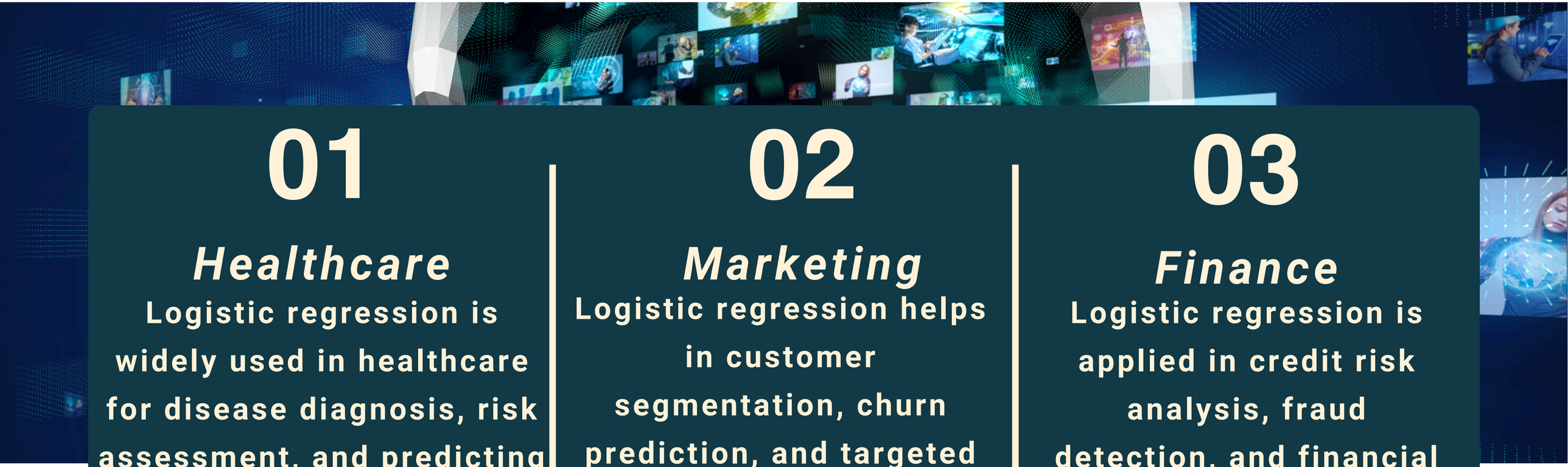
04

Fraud Detection

Logistic regression is used to detect fraudulent activities in areas such as insurance claims, financial transactions, and online purchases.

03 USE CASES

Logistic Regression is used in three cases: Healthcare, Marketing and Finance



01

Healthcare

Logistic regression is widely used in healthcare for disease diagnosis, risk assessment, and predicting treatment outcomes.

02

Marketing

Logistic regression helps in customer segmentation, churn prediction, and targeted marketing campaigns.

03

Finance

Logistic regression is applied in credit risk analysis, fraud detection, and financial risk management.

04

DATASET

We'll be trying to predict a classification- survival or deceased. Let's begin our understanding of implementing Logistic Regression in Python for classification.

Passenger	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, M	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cumings, I	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen,	female	26	0	0	STON/O2.	7.925		S
4	1	1	Futrelle, M	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr.	male	35	0	0	373450	8.05		S
6	0	3	Moran, M	male		0	0	330877	8.4583		Q
7	0	1	McCarthy,	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, M	male	2	3	1	349909	21.075		S
9	1	3	Johnson, I	female	27	0	2	347742	11.1333		S
10	1	2	Nasser, M	female	14	1	0	237736	30.0708		C
11	1	3	Sandstron	female	4	1	1	PP 9549	16.7	G6	S
12	1	1	Bonnell, M	female	58	0	0	113783	26.55	C103	S
13	0	3	Saundercc	male	20	0	0	A/5. 2151	8.05		S
14	0	3	Anderssor	male	39	1	5	347082	31.275		S
15	0	3	Vestrom, I	female	14	0	0	350406	7.8542		S
16	1	2	Hewlett, M	female	55	0	0	248706	16		S
17	0	3	Rice, Mast	male	2	4	1	382652	29.125		Q
18	1	2	Williams, I	male		0	0	244373	13		S

```
[ ]  from sklearn.preprocessing import *
     from sklearn.model_selection import *                #for splitting
     from sklearn.linear_model import LogisticRegression
     from sklearn.metrics import *
     from sklearn.model_selection import train_test_split
     from sklearn.metrics import accuracy_score
```

```
     # Split data into features (X) and target (y)
     X = data.drop(['Survived', 'PassengerId', 'Name', 'Ticket'], axis=1)
     y = data['Survived']
```

```
[ ]  # Split data into train and test sets
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
[ ]  X_train
```

```
[ ]  # Create and fit the logistic regression model
     model = LogisticRegression()
```

```python
# Create and fit the logistic regression model
model = LogisticRegression()
model.fit(X_train, y_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:458: ConvergenceWarning: lbfgs fai
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

▾ LogisticRegression

LogisticRegression()

```python
# Make predictions on the test set
y_pred = model.predict(X_test)
```

+ Code + Text

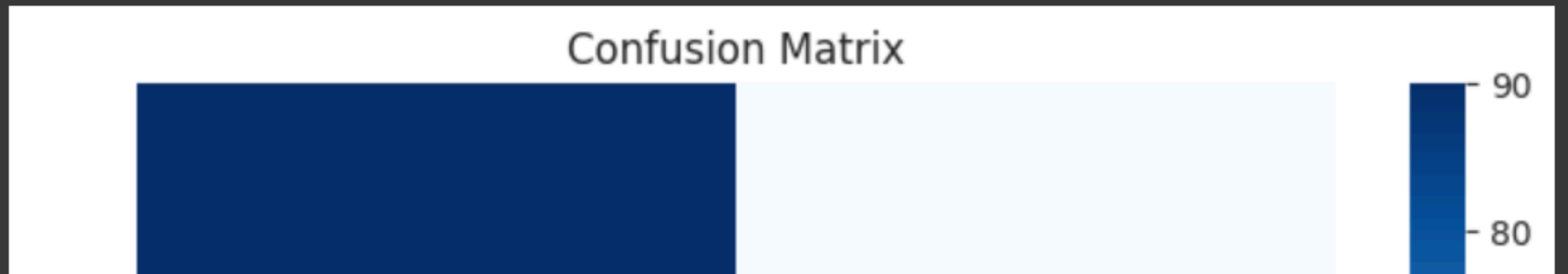
```python
y_pred
```

```
array([0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1,
```

```
[ ]  # Confusion matrix
     conf_matrix = confusion_matrix(y_test, y_pred)
     print("Confusion Matrix:")
     print(conf_matrix)
```

```
Confusion Matrix:
[[90 15]
 [21 53]]
```

```
[ ]  # Plot the confusion matrix as a heatmap
     plt.figure(figsize=(8, 6))
     sns.heatmap(conf_matrix, annot=True, cmap='Blues', fmt='d')
     plt.title('Confusion Matrix')
     plt.xlabel('Predicted Values')
     plt.ylabel('Actual Values')
     plt.show()
```



CONCLUSION

In conclusion, this analysis applied a logistic regression model to predict passenger survival on the Titanic using features from the available dataset. The model achieved an accuracy of 79%, correctly classifying the survival status for nearly 4 out of every 5 passengers.

Any questions?

THANK YOU!!