

VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY  
HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY  
FACULTY OF COMPUTER SCIENCE AND ENGINEERING



## Probability and Statistics

---

### Project Report

# Predictive Modeling of Internet Advertisement Characteristics

---

Instructor(s): Nguyen Tien Dung

Student(s): Pham Viet Hao 2352297

Nguyen Minh Hoang 2352352

Lu Thuan Hung 2352429

Thuong Dinh Hung 2352444

Huynh Nhat Huy 2352381

HO CHI MINH CITY, APRIL 2025





## Member list & Workload

No.	Full name	Student ID	Contribution
1	Pham Viet Hao	2352297	20%
2	Nguyen Minh Hoang	2352352	20%
3	Lu Thuan Hung	2352429	20%
4	Thuong Dinh Hung	2352444	20%
5	Huynh Nhat Huy	2352381	20%

## Contents

<b>1</b>	<b>Project Overview</b>	<b>4</b>
1.1	Project Introduction . . . . .	4
1.2	Data Context . . . . .	4
1.3	Monitoring Values . . . . .	4
1.4	Variables . . . . .	4
<b>2</b>	<b>Background Knowledge</b>	<b>4</b>
<b>3</b>	<b>Data Preprocessing</b>	<b>4</b>
3.1	Import Libraries . . . . .	4
3.2	Import Data . . . . .	4
3.3	Reading Data . . . . .	5
<b>4</b>	<b>Descriptive Statistics</b>	<b>5</b>
<b>5</b>	<b>Inferential Statistics</b>	<b>5</b>
<b>6</b>	<b>Discussion and Extensions</b>	<b>5</b>
<b>7</b>	<b>Data Sources and Code Source</b>	<b>5</b>



# 1 Project Overview

## 1.1 Project Introduction

In today's digital landscape, understanding and be able to classify advertisements is crucial for various applications such as ad blocking software, market analytics, etc. Therefore, in this project, we aim to build a predictive model that can classify advertisements based attributes such as height, width, and aspect ratio.

Source: [www.kaggle.com/datasets/uciml/internet-advertisements-data-set/data](https://www.kaggle.com/datasets/uciml/internet-advertisements-data-set/data)

## 1.2 Data Context

The data used in this project is given by UCI Machine Learning Respository, contains 1558 columns and 3278 rows with each row represents one image tagged as ad or nonad in the last column. The actual numerical attributes of the images are represented from column 0 to 1557 with 3 continuous attributes.

## 1.3 Monitoring Values

## 1.4 Variables

Variable	Type	Description
0	Numeric	Height of image
1	Numeric	Width of image
2	Numeric	Aspect ratio of image
3-1557	Numeric	Image attributes
1558	Character	Categoriize between ad and nonad

# 2 Background Knowledge

just change the branch 2

# 3 Data Preprocessing

## 3.1 Import Libraries

## 3.2 Import Data



```
1 data <- read.csv("add.csv")
```

### 3.3 Reading Data

## 4 Descriptive Statistics

## 5 Inferential Statistics

## 6 Discussion and Extensions

## 7 Data Sources and Code Source

```
1 cout << "hello"  
2 printf("")  
3     print("nested")
```

```
C:\Users\Group6>R
```

```
R version 4.5.0 (2025-04-11 ucrt) -- "How About a Twenty-Six"  
Copyright (C) 2025 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64
```

```
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.
```

```
  Natural language support but running in an English locale
```

```
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.
```

```
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```



>

```
1 data <- read.csv("add.csv")
2 print(dim(data))
3
4 # cars <- c(1, 3, 6, 4, 9)
5 # trucks <- c(2, 5, 4, 5, 12)
6
7 # plot(cars, type="o", col="blue", ylim=c(0,12))
8
9 # lines(trucks, type="o", pch=22, lty=2, col="red")
10
11 # title(main="Autos", col.main="red", font.main=4)
```

```
1 data <- read.csv("add.csv")
2 print(dim(data))
```

## References

- [1] M. Lichman. UCI machine learning repository, 2013.